



ulm university universität
uulm

Ökonometrie

Vorlesungsskript

Prof. Dr. Evgeny Spodarev

Ulm Wintersemester 2008/2009

Vorwort

Dieses Skript entstand aus dem Zyklus der Vorlesungen über Ökonometrie, die ich in den Jahren 2008-2009 an der Universität Ulm gehalten habe. Dabei handelt es sich um die erste Einführung in die Ökonometrie, die auf Master-Niveau für die Studenten der Wirtschaftswissenschaften angeboten wird.

Ich möchte gerne Herrn Dipl. Math. oec. Wolfgang Karcher, der die Übungen zu dieser Vorlesung gehalten hat, für seine Unterstützung während der Entstehung des Skriptes danken. Herr Gregor Hesse hat eine hervorragende Arbeit beim Tippen des Skriptes und bei der Erstellung zahlreicher Abbildungen, die den Text begleiten, geleistet. Dafür gilt ihm mein herzlicher Dank.

Ulm, den 09.03.2009

Evgeny Spodarev

Inhaltsverzeichnis

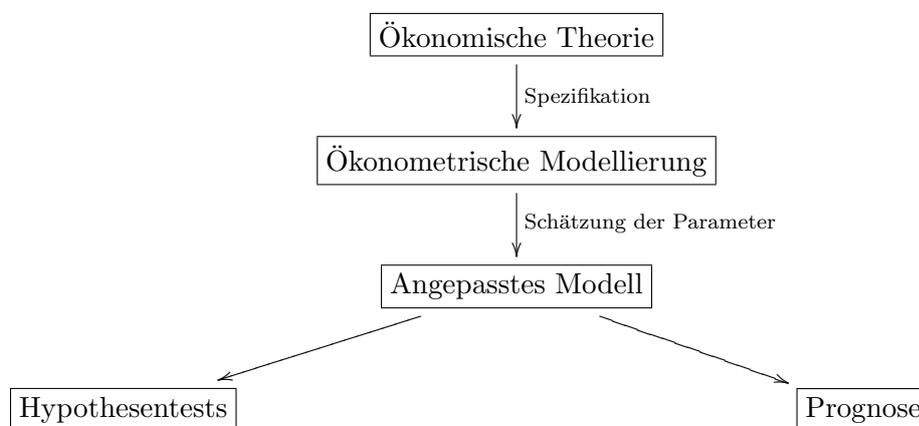
1	Ökonomische Modellbildung	5
1.1	Was ist Ökonometrie?	5
1.2	Ökonometrisches Modell	6
1.3	Geschichtliche Entwicklung der Ökonometrie	7
2	Verallgemeinerte lineare Modelle	7
2.1	Warum braucht man nichtlineare Modelle?	7
2.2	Rettungsversuch für das lineare Modell	8
2.2.1	Quasilineare Modelle	8
2.2.2	Lineare Approximation	10
2.3	Informelle Einführung der verallgemeinerten linearen Modelle	12
2.4	Nichtlineare Zusammenhänge in den Zielvariablen	16
2.4.1	Maximum-Likelihood-Methode	18
2.4.2	Box-Cox-Transformation	21
2.5	Test der Linearität	26
2.5.1	Reset-Tests	26
2.5.2	Welches Modell ist besser? - Bestimmtheitsmaß R^2	30
2.5.3	Box-Cox-Verfahren	31
2.5.4	Likelihood-Ratio-Test	34
2.6	Verallgemeinerte lineare Modelle	36
2.6.1	Maximum-Likelihood-Schätzung von β	41
2.6.2	Asymptotische Tests für β	47
2.6.3	Kriterien zur Modellwahl bzw. Modellanpassung	53
3	Zeitreihenanalyse	57
3.1	Korrelierte Beobachtungen im Regressionsmodell	57
3.2	Zeitreihen	58

3.2.1	Autoregressive Prozesse: eine Einführung	59
3.2.2	Typen von Zeitreihen	62
3.2.3	Vorhersage von Zeitreihen	70
3.2.4	Eigenschaften der Zeitreihen	73
3.3	Statistik in der Zeitreihenanalyse	79
3.3.1	Schätzung der Kovarianzfunktion	79
3.3.2	Parameterschätzung bei AR(p)-Prozessen	80
4	Appendix	83
4.1	Regression	83
4.1.1	Einfache lineare Regression	83
4.1.2	Multivariate lineare Regression	85
4.2	p-Wert	86
	Literaturverzeichnis	89
	Index	90

1 Ökonomische Modellbildung

1.1 Was ist Ökonometrie?

Ökonometrie ist ein Teilgebiet der Wirtschaftswissenschaften, das sich mit der Anwendung der mathematischen Statistik auf die quantitative Erklärung von empirischen Zusammenhängen in der Wirtschaft beschäftigt. Sie analysiert anhand von beobachtbaren Daten ökonomische Wirkungszusammenhänge, die hinter den Daten vermutet werden. Dabei bedient sie sich einerseits der ökonomischen Theorie, andererseits der Methoden der Mathematik und insbesondere der Stochastik. Die Arbeitsweise eines Ökonometrikers wird hierfür im folgenden Schaubild angedeutet:



Dabei werden die relevanten Züge eines wirtschaftlichen Phänomens durch das Abstrahieren in eine *ökonomische Theorie* überführt, die als Basis für die Entwicklung eines quantitativen *ökonomischen Modells* dient. Ein ökonometrisches Modell beinhaltet eine in der Regel relativ kleine Anzahl von Parametern, die aus den vorhandenen Wirtschaftsdaten geschätzt werden sollten, um eine möglichst gute Anpassung des Modells an die Daten zu gewährleisten. Das auf diese Weise angepasste Modell dient der Erstellung und Durchführung von Hypothesentests und ökonometrischen Prognosen.

Welche Arten von Daten stehen einem Ökonometriker zur Verfügung?

Es sind

- Zeitreihen
- Querschnittsdaten
- Paneldaten

Zeitreihen entstehen durch die Beobachtung eines einzelnen Wirtschaftssubjekts zu mehreren Zeitpunkten (z.B. über mehrere Jahre).

Querschnittsdaten entstehen durch die gleichzeitige einmalige Beobachtung mehrerer Wirtschaftssubjekte.

Paneldaten entstehen durch die Beobachtung mehrerer Wirtschaftssubjekte zu mehreren Zeitpunkten. Damit stellen Paneldaten eine Kombination aus Zeitreihen- und Querschnittsdaten dar.

In dieser Vorlesung beschäftigen wir uns hauptsächlich mit Querschnittsdaten und Zeitreihen, seltener dagegen mit Paneldaten.

1.2 Ökonometrisches Modell

Eines der Hauptmodelle der Ökonometrie kann folgendermaßen formuliert werden:

$$Y_i = f(x_{i1}, \dots, x_{im}) + \varepsilon_i, \quad i = 1, \dots, n \quad (1.2.1)$$

Dabei sind Y_i die sog. unabhängigen Variablen, die oft als *Regressanden* oder *endogene Variablen* bezeichnet werden. Sie sind oft als mehrere Beobachtungen eines wirtschaftlichen Merkmals Y zu interpretieren.

Die Variablen x_{i1}, \dots, x_{im} , $i = 1, \dots, n$ sind Ausgangsvariablen, die auch *Regressoren* oder *exogene Variablen* genannt werden. Die Variablen ε_i sind Störgrößen, die z.B. durch Beobachtungs- oder Messfehler entstehen können. Sie beinhalten aber auch quantitative Einflüsse, die in (1.2.1) nicht berücksichtigt wurden und auf zufällige Schwankungen zurückzuführen sind. In der Vorlesung „*Stochastik für Wirtschaftswissenschaftler*“ und „*Wirtschaftsstatistik*“ wurde die Situation der linearen Regression betrachtet, in der

$$f(x_1, \dots, x_m) = \beta_1 x_1 + \dots + \beta_m x_m$$

eine lineare Funktion ist, x_{ij} deterministisch und ε_i unkorreliert (oder unabhängig) mit der selben Varianz σ^2 sind. Oft wurde dabei die Annahme der Normalverteilung von ε_i benutzt. In dieser vertiefenden Vorlesung „*Ökonometrie*“ werden allgemeinere Funktionen f zugelassen, die zu flexiblen Modellierungsmöglichkeiten führen. Solche Modelle heißen *verallgemeinerte lineare Modelle*. Sie werden in Kapitel 2 untersucht. Außerdem wird die Annahme der Unkorreliertheit von ε_i abgeschwächt, was zu den Begriffen der Zeitreihen und der stationären stochastischen Prozesse führt. Sie werden in Kapitel 3 behandelt. Die Vorlesung wird durch zahlreiche Zahlenbeispiele belebt, die Anwendungen ökonometrischer Theorie in der Wirtschaftspraxis aufzeigen.

1.3 Geschichtliche Entwicklung der Ökonometrie

Die ersten systematischen statistischen Studien in der Ökonometrie gehen auf das Jahr 1914 zurück, in dem H. L. Moore die Nachfrage statistisch untersucht hat. E. J. Working (1926) und H. Schultz (1938) haben diese Richtung bis zur Monographie "The theory and measurement of demand" weiterentwickelt. So wurde in den 1940er Jahren durch die Arbeiten von H. Mann, A. Wald, T. Haavelmo und anderer die Basis der Ökonometrie gelegt, die in den 1970er Jahren ihre moderne Form annahm. Seitdem wurden ihre Methoden stets weiterentwickelt. Heute bedient sich die Ökonometrie zahlreicher Methoden aus unterschiedlichen Bereichen der Stochastik.



H.L. Moore
(1869 - 1958)

2 Verallgemeinerte lineare Modelle

2.1 Warum braucht man nichtlineare Modelle?

In den meisten ökonometrischen Anwendungen sind die beobachtbaren Zusammenhänge essentiell nichtlinear.

Beispiel 2.1.1 (Warenherstellung)

Ein linearer Zuwachs eines Produktionsparameters (Input) beim gleichzeitigen Konstanthalten aller anderen Produktionsparameter führt in der Regel nicht zum selben linearen Zuwachs der Produktion, sondern ist meist kleiner.

Beispiel 2.1.2 (Milchproduktion)

vgl. v. Auer - Ökonometrie, S. 285ff

In folgender Tabelle wird der Zusammenhang zwischen der jährlichen Kraftfutteraufnahme einer Kuh und deren jährlicher Milchproduktion dargestellt.

$Y_i, i = 1, \dots, 12$ - Milchleistung der Kuh i in l/Jahr

$x_i, i = 1, \dots, 12$ - Kraftfutterabgabe in Zentner/Jahr

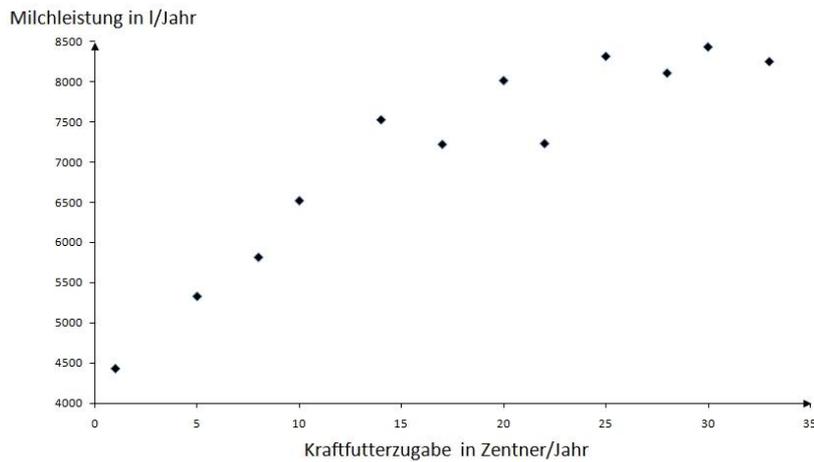
	1	2	3	4	5	6	7	8	9	10	11	12
Y_i	6525	8437	8019	8255	5335	7236	5821	7531	8320	4436	7225	8112
x_i	10	30	20	33	5	22	8	14	25	1	17	28

Tabelle 2.1.2

Einfache lineare Regression:

$$Y_i = \alpha + \beta x_i + \varepsilon_i, \quad i = 1, \dots, 12$$

Plottet man die Punktwolke (x_i, Y_i) auf der xy -Ebene, so wird aus der Graphik ersichtlich, dass die Linearitätsannahme nicht erfüllt ist.



2.2 Rettungsversuch für das lineare Modell

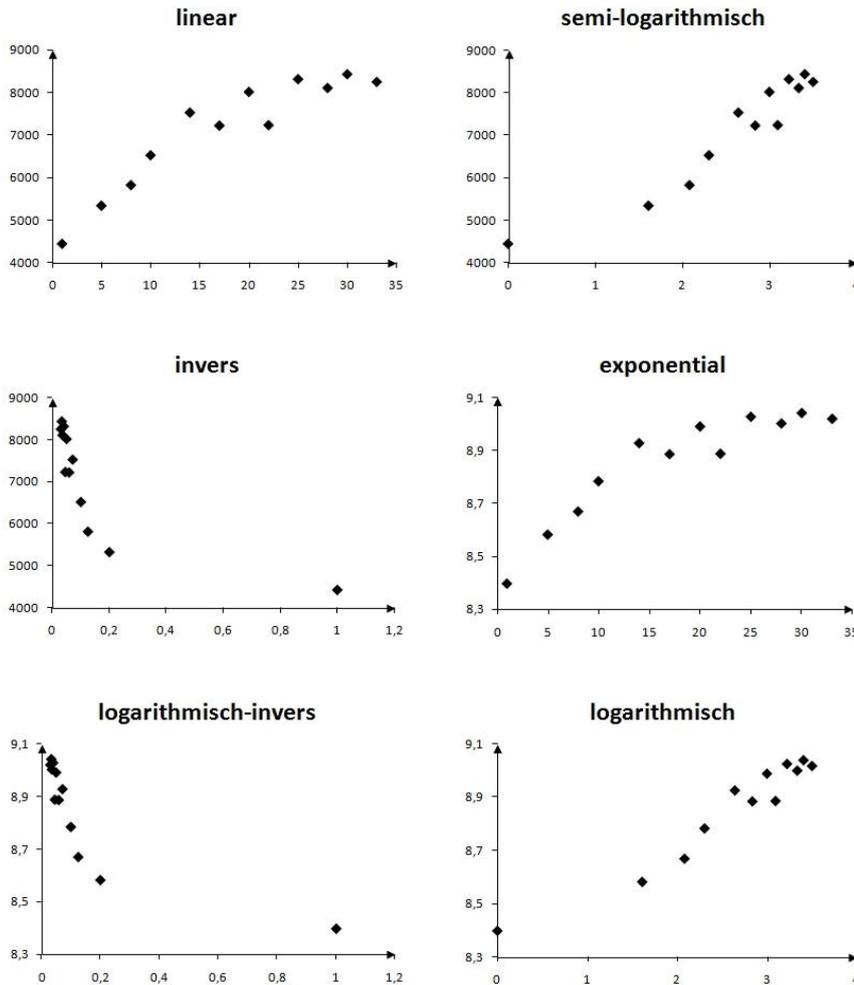
2.2.1 Quasilineare Modelle

Wir geben mehrere Beispiele für nichtlineare funktionale Zusammenhänge an, die in der Ökonometrie verwendet werden.

Name	Funktionstyp
linear	$Y_i = \alpha + \beta x_i (+ \varepsilon_i)$
semi-logarithmisch	$Y_i = \alpha + \beta \log x_i (+ \varepsilon_i)$
invers	$Y_i = \alpha + \frac{\beta}{x_i} (+ \varepsilon_i)$
exponential	$\log Y_i = \alpha + \beta x_i (+ \varepsilon_i)$
logarithmisch	$\log Y_i = \alpha + \beta \log x_i (+ \varepsilon_i)$
logarithmisch-invers	$\log Y_i = \alpha + \frac{\beta}{x_i} (+ \varepsilon_i)$
quadratisch	$Y_i = \alpha + \beta_2 x_i + \beta_3 x_i^2 (+ \varepsilon_i)$

Tabelle 2.2.1

Für Beispiel 2.1.2 zeichnen wir die entsprechenden Punktwolken der Modelle aus der Tabelle 2.2.1 auf unterschiedlichen Skalen (z.B. lineare, logarithmische, exponentielle oder inverse Skalierung) und versuchen die Linearität in den Punktwolken festzustellen.



Das logarithmische Modell passt am besten zu den Daten.

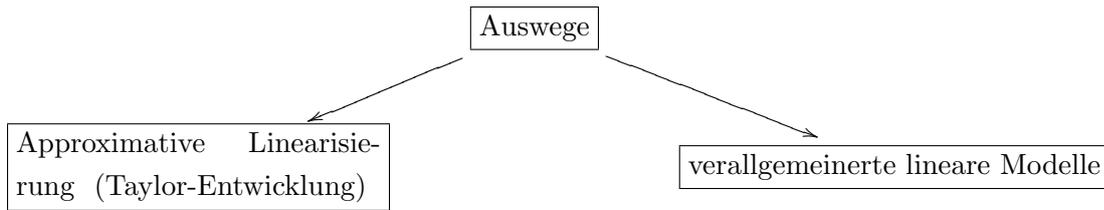
$$\underbrace{\log Y_i}_{Z_i} = \alpha + \beta \underbrace{\log x_i}_{t_i} + (\varepsilon_i) \Leftrightarrow Y_i = e^\alpha x_i^\beta e^{\varepsilon_i}, \quad i = 1, \dots, 12$$

Falls jedoch die neuen Variablen $Z_i = \log Y_i$ und $t_i = \log x_i$ eingeführt werden, wird das Modell $Z_i = \alpha + \beta t_i + \varepsilon_i$ linear und es kann die Standardanalyse für die einfache lineare Regression mit den Variablen Z_i, t_i durchgeführt werden. Danach erfolgt die Umrechnung in Y_i und x_i .

Beispiel 2.2.1

Es gibt aber auch Beispiele von funktionalen Zusammenhängen, die nicht so einfach (bzw. gar nicht) linearisiert werden können:

$$Y_i = \beta_1 + \beta_2 e^{\beta_3 x_i} + \varepsilon_i, \quad i = 1, \dots, n$$



2.2.2 Lineare Approximation

Ein essentiell nichtlineares (und kein quasilineares) Modell

$$Y_i = f(x_{i1}, \dots, x_{il}, \beta_1, \dots, \beta_m) + \varepsilon_i \stackrel{\text{Kurzschreibweise}}{=} f(\vec{x}_i, \beta) + \varepsilon_i, \quad i = 1, \dots, n$$

$$\text{mit } \vec{x}_i = (x_{i1}, \dots, x_{il}), \quad i = 1, \dots, n, \quad \beta = (\beta_1, \dots, \beta_m), \quad f: \mathbb{R}^{l \times m} \rightarrow \mathbb{R}$$

kann in manchen Fällen dennoch mit Methoden der linearen Regression untersucht werden. Dabei wird f (unter Annahme der zweimaligen Differenzierbarkeit) in die Taylor-Reihe bzgl. der Variablen $\beta \in \mathbb{R}^m$ im Punkt $\beta^0 \in \mathbb{R}^m$ entwickelt:

$$f(\vec{x}, \beta) = f(\vec{x}, \beta^0) + \sum_{k=1}^m \frac{\partial f(\vec{x}, \beta)}{\partial \beta_k} \Big|_{\beta=\beta^0} (\beta_k - \beta_k^0) + o(|\beta - \beta^0|), \quad (2.2.1)$$

falls β in einer kleinen Umgebung von β^0 liegt. Der asymptotisch kleine Term $o(|\beta - \beta^0|)$ (für $\beta \rightarrow \beta^0$) wird bei der weiteren Analyse vernachlässigt.

Schreiben wir die Darstellung (2.2.1) folgendermaßen um:

$$f(\vec{x}, \beta) \approx f(\vec{x}, \beta^0) + \sum_{k=1}^m \beta_k \frac{\partial f(\vec{x}, \beta)}{\partial \beta_k} \Big|_{\beta=\beta^0} - \sum_{k=1}^m \beta_k^0 \frac{\partial f(\vec{x}, \beta)}{\partial \beta_k} \Big|_{\beta=\beta^0}$$

Führen wir folgende abkürzende Bezeichnungen ein:

$$\begin{aligned} \tilde{x}_k &= \frac{\partial f(\vec{x}, \beta)}{\partial \beta_k}, \quad k = 1, \dots, m & \tilde{x}_k^0 &= \frac{\partial f(\vec{x}, \beta)}{\partial \beta_k} \Big|_{\beta=\beta^0} \\ \tilde{x}_{i,k} &= \frac{\partial f(\vec{x}_i, \beta)}{\partial \beta_k}, \quad i = 1, \dots, n & \tilde{x}_{i,k}^0 &= \frac{\partial f(\vec{x}_i, \beta)}{\partial \beta_k} \Big|_{\beta=\beta^0} \\ \tilde{x} &= (\tilde{x}_1, \dots, \tilde{x}_m)^\top & \tilde{x}^0 &= (\tilde{x}_1^0, \dots, \tilde{x}_m^0)^\top \\ \tilde{X} &= (\tilde{x}_{i,k})_{\substack{i=1, \dots, n \\ k=1, \dots, m}} & \tilde{X}^0 &= (\tilde{x}_{i,k}^0)_{\substack{i=1, \dots, n \\ k=1, \dots, m}} \end{aligned}$$

Dann gilt:

$$f(\vec{x}, \beta) \approx f(\vec{x}, \beta^0) + (\tilde{x}^0)^\top \beta - (\tilde{x}^0)^\top \beta^0$$

oder

$$Y_i \approx f(\vec{x}_i, \beta^0) + (\tilde{x}^0)_i^\top \beta + \varepsilon_i - (\tilde{x}^0)_i^\top \beta^0, \quad i = 1, \dots, n,$$

wobei durch $(\tilde{x}^0)_i$ die i -te Zeile der Matrix \tilde{X}^0 bezeichnet wird.

Transformieren wir die unabhängigen Variablen Y_i nach der Regel

$$\tilde{Y}_i^0 = Y_i - f(\vec{x}_i, \beta^0) + (\tilde{x}^0)_i^\top \beta^0,$$

so bekommen wir ein neues lineares Modell

$$\tilde{Y}_i^0 \approx (\tilde{x}^0)_i^\top \beta + \varepsilon_i, \quad i = 1, \dots, n, \quad (2.2.2)$$

das als Approximation des ursprünglichen Modells angesehen werden kann. Dieses neue lineare Modell kann mit den üblichen Mitteln für lineare Modelle analysiert werden (MKQ-Methode, usw.). Falls β^0 als erste Näherung an β gegeben ist (z.B. aus Überlegungen der ökonomischen Theorie), dann kann ein Schätzer $\hat{\beta}$ für β wie im Abschnitt 4.1.1 des Appendix gewonnen werden.

Beispiel 2.2.2

Betrachten wir das nichtlineare Zusammenhangsmodell aus Beispiel 2.1.2 :

$$f(x, \beta) = \beta_1 + \beta_2 e^{\beta_3 x}, \quad x \in \mathbb{R}$$

Somit ist $l=1$, $m=3$ und für das nichtlineare Modell

$$Y_i = f(x_i, \beta) + \varepsilon_i, \quad i = 1, \dots, n$$

kann das linearisierte Modell folgendermaßen konstruiert werden:

Für ein $\beta^0 = (\beta_1^0, \beta_2^0, \beta_3^0)^\top$ berechnen wir zunächst die partiellen Ableitungen:

$$\begin{aligned}\tilde{x}_1^0 &= \left. \frac{\partial f(x, \beta)}{\partial \beta_1} \right|_{\beta=\beta_0} = 1, \\ \tilde{x}_2^0 &= \left. \frac{\partial f(x, \beta)}{\partial \beta_2} \right|_{\beta=\beta_0} = e^{\beta_3^0 x}, \\ \tilde{x}_3^0 &= \left. \frac{\partial f(x, \beta)}{\partial \beta_3} \right|_{\beta=\beta_0} = \beta_2^0 x e^{\beta_3^0 x}.\end{aligned}$$

Somit ist die lineare Approximation durch

$$\begin{aligned}\tilde{Y}_i^0 &= Y_i - f(x_i, \beta^0) + \sum_{k=1}^3 \beta_k^0 (\tilde{x}_k^0)_i \\ &= Y_i - \beta_1^0 - \beta_2^0 e^{\beta_3^0 x_i} + \beta_1^0 + \beta_2^0 e^{\beta_3^0 x_i} + \beta_3^0 \beta_2^0 x_i e^{\beta_3^0 x_i}, \\ \tilde{Y}_i^0 &\approx \beta_1 + \beta_2 e^{\beta_3 x_i} + \beta_3 \beta_2 x_i e^{\beta_3 x_i} + \varepsilon_i, \quad i = 1, \dots, n\end{aligned}$$

gegeben.

2.3 Informelle Einführung der verallgemeinerten linearen Modelle

Wie wird aber im allgemeinen Fall beim nichtlinearen Modell

$$Y_i = f(\vec{x}_i, \beta) + \varepsilon_i, \quad i = 1, \dots, n$$

vorgegangen, um die Parameter β und $\sigma^2 = \varepsilon_i$ des Modells zu schätzen? Nach wie vor setzen wir voraus, dass die Störgrößen unkorreliert sind mit

$$\mathbb{E} \varepsilon_i = 0 \text{ und } \text{Var} \varepsilon_i = \sigma^2, \quad i = 1, \dots, n.$$

Auch hier kann β durch die MKQ-Methode geschätzt werden:

Der mittlere quadratische Fehler

$$e(\beta) = \frac{1}{n} \sum_{i=1}^n (Y_i - f(\vec{x}_i, \beta))^2 = \frac{1}{n} \sum_{i=1}^n \varepsilon_i^2$$

wird gebildet und bezüglich β minimiert:

$$e(\beta) \rightarrow \min_{\beta}$$

Nun ist es aber ein (im Allgemeinen) nicht-quadratisches Optimierungsproblem, das zu den notwendigen Bedingungen des Extremums

$$\frac{\partial e(\beta)}{\partial \beta_k} = 0, \quad k = 1, \dots, m$$

führt. Da

$$\frac{\partial e(\beta)}{\partial \beta_k} = -\frac{2}{n} \sum_{i=1}^n (Y_i - f(\vec{x}_i, \beta)) \frac{\partial f(\vec{x}_i, \beta)}{\partial \beta_k}$$

ist, erhält man daraus ein System von m nichtlinearen Gleichungen

$$\sum_{i=1}^n (Y_i - f(\vec{x}_i, \beta)) \frac{\partial f(\vec{x}_i, \beta)}{\partial \beta_k} = 0, \quad k = 1, \dots, m$$

oder in gekürzter Matrix-Form

$$\tilde{X}^\top \varepsilon = 0$$

mit $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)^\top$ (Analog zur Normalengleichung im linearen Fall).

Problem:

Das Normalengleichungssystem ist nichtlinear, deswegen braucht man numerische Methoden, um es bzgl. β aufzulösen.

Unter der Voraussetzung, dass die Lösung des Systems (der MKQ-Schätzer $\hat{\beta}$) existiert, eindeutig ist und einen Minimum-Punkt darstellt, können numerische Methoden zur Berechnung von $\hat{\beta}$ eingesetzt werden, die wir etwas später detaillierter behandeln werden. Hier sei nur das einfache *Gauß-Newton-Verfahren* erwähnt, dass in enger Verbindung mit den linearisierten Modellen steht.

Iterativer Algorithmus zur Suche von $\hat{\beta}$:

1. Initialisierung:

Fixiere einen Wert β^0 und linearisiere das oben genannte Modell mit Hilfe der Taylor-Entwicklung wie in Abschnitt 2.2.2 beschrieben. Führe die MKQ-Methode zur Suche von Schätzern für β im linearisierten Modell durch. Das Ergebnis der Schätzung sei ein Parameter-Vektor $\beta^1 = (\beta_1^1, \dots, \beta_m^1)^\top$.

2. Iterationsschritt:

Sei $\beta^k = (\beta_1^k, \dots, \beta_m^k)^\top$ das Ergebnis des Iterationsverfahrens im letzten (k -ten) Schritt,

$k \in \mathbb{N}$. Benutze β^k als Startwert β^0 bei der Linearisierung des Modells (2.2.2) und berechne den MKQ-Schätzer für β in (2.2.2). Das Ergebnis wird mit $\beta^{k+1} = (\beta_1^{k+1}, \dots, \beta_m^{k+1})^\top$ bezeichnet.

3. Abbruchregel:

Iteriere das o.g. Verfahren solange bis zum ersten Mal $|\beta^{k+1} - \beta^k| \leq \delta$ gilt, wobei $\delta > 0$ eine vorgegebene Präzision ist.

4. Endergebnis:

Wähle das erste β^{k+1} , so dass $|\beta^{k+1} - \beta^k| \leq \delta$ gilt.

Bemerkung 2.3.1

Ein schwer überwindbarer Nachteil des Verfahrens liegt in der Auswahl der Initialisierung β^0 . Sie sollte möglichst nah an dem tatsächlichen Wert von β liegen, damit das Verfahren konvergiert. Sehr oft ist aber so ein Wert in der Praxis nicht bekannt. Man sollte durch mehrfaches Ausprobieren ein Gefühl dafür bekommen, welche Werte für β^0 in Frage kommen. Die soeben erläuterte Vorgehensweise soll im folgenden numerischen Beispiel illustriert werden:

Beispiel 2.3.1

vgl. Greene - Econometric Analysis, S. 318f

Das Jahreseinkommen und der Jahreskonsum (in Milliarden US-Dollar) der amerikanischen Haushalte in den Jahren von 1950 bis 1985 sind in folgender Tabelle dargestellt.

- X_i = Jahreseinkommen im Jahr i
- Y_i = Jahreskonsum im Jahr i

	1950	1951	1952	1953	1954	1955	1956	1957	1958
X_i	791.8	819.0	844.3	880.0	894.0	944.5	989.4	1012.1	1028.8
Y_i	733.2	748.7	771.4	802.5	822.7	973.8	899.8	919.7	932.9
	1959	1960	1961	1962	1963	1964	1965	1966	1967
X_i	1067.2	1091.1	1123.2	1170.2	1207.3	1291.0	1365.7	1431.3	1493.2
Y_i	979.4	1005.1	1025.2	1069.0	1108.4	1170.6	1236.4	1298.9	1337.7
	1968	1969	1970	1971	1972	1973	1974	1975	1976
X_i	1551.3	1599.8	1688.1	1728.4	1797.4	1916.3	1896.6	1931.7	2001.0
Y_i	1405.9	1456.7	1492.0	1538.8	1621.9	1689.6	1674.0	1711.9	1803.9
	1977	1978	1979	1980	1981	1982	1983	1984	1985
X_i	2066.6	2167.4	2212.6	2214.3	2248.6	2261.5	2334.6	2468.4	2509.0
Y_i	1883.8	1961.0	2004.4	2000.4	2024.2	2050.7	2145.9	2239.9	2312.6

Tabelle 2.3.1

Es wird das folgende nichtlineare Modell für den Zusammenhang der beiden Größen unterstellt:

$$Y_i = \beta_1 + \beta_2 x_i^{\beta_3} + \varepsilon_i, \quad i = 1, \dots, n \quad (n = 35)$$

Die Funktion

$$f(x_i, \beta) = \beta_1 + \beta_2 x_i^{\beta_3},$$

mit $\beta_3 \geq 0$, $\beta = (\beta_1, \beta_2, \beta_3)^\top$, heißt dabei *Konsumfunktion*. Es wird vermutet, dass β_3 den Wert 1 leicht übersteigt (der Fall $\beta_3 = 1$ entspricht dem linearen Modell). Als erstes wird das Modell linearisiert:

$$\begin{aligned} \frac{\partial f(x_i, \beta)}{\partial \beta_1} \Big|_{\beta=\beta_0} &= 1, \\ \frac{\partial f(x_i, \beta)}{\partial \beta_2} \Big|_{\beta=\beta_0} &= x_i^{\beta_3^0}, \\ \frac{\partial f(x_i, \beta)}{\partial \beta_3} \Big|_{\beta=\beta_0} &= \beta_2^0 x_i^{\beta_3^0} \log x_i. \end{aligned}$$

$$\begin{aligned}
\tilde{Y}_i^0 &= Y_i - f(x_i, \beta^0) + \sum_{k=1}^m \frac{\partial f(x_i, \beta)}{\partial \beta_k} \beta_k^0 \\
&= Y_i - \beta_1^0 - \beta_2^0 x_i^{\beta_3^0} + \beta_1^0 + \beta_2^0 x_i^{\beta_3^0} + \beta_3^0 \beta_2^0 x_i^{\beta_3^0} \log x_i = Y_i + \beta_3^0 \beta_2^0 x_i^{\beta_3^0} \log x_i, \\
\tilde{Y}_i^0 &\approx \beta_1 + \beta_2 x_i^{\beta_3} + \beta_3 \beta_2 x_i^{\beta_3} \log x_i + \varepsilon_i
\end{aligned}$$

Dabei wird für β^0 das Ergebnis des MKQ-Schätzers von β im linearen Modell ($\beta_3 = 1$) genommen:

$$\beta_1^0 = 11.1458, \quad \beta_2^0 = 0.848534, \quad \beta_3^0 = 1$$

It. - Nr.	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
1	11.1458	0.898534	1
2	209.825	-0.237125	1.15139
3	187.786	0.246078	1.14613
4	187.710	0.245692	1.15699
5	187.915	0.245968	1.15641
6	187.899	0.246004	1.15640

Tabelle 2.3.2 - Ergebnisse der ersten 6 Iterationen

Man sieht leicht, dass die Werte schnell konvergieren und zu folgendem Ergebnis führen:

$$\hat{\beta}_1 = 187.899, \quad \hat{\beta}_2 = 0.246004, \quad \hat{\beta}_3 = 1.15640 > 1$$

2.4 Nichtlineare Zusammenhänge in den Zielvariablen

Manchmal ist es vorteilhaft, Regressionszusammenhänge der Form

$$g(Y_i, \theta) = f(\vec{x}_i, \beta) + \varepsilon_i, \quad i = 1, \dots, n \quad (2.4.1)$$

zu betrachten, die auch auf der linken Seite der Regressionsgleichung eine Nichtlinearität bzgl. der Zielvariablen Y_i aufweisen. Dabei hängt $g(Y_i, \theta)$ nicht nur von Y_i , sondern auch von dem neuen Parametervektor $\theta = (\theta_1, \dots, \theta_k)$, $k \geq 1$ ab, der (zusammen mit β) geschätzt werden soll.

Beispiel 2.4.1 (Verallgemeinerte Produktionsfunktion)

vgl. *Greene - Econometric Analysis, S. 327ff*

In den Produktionsuntersuchungen wird folgender Zusammenhang zwischen dem Produktionsvolumen Y_i , dem Kapital x_{i1} und dem Personalaufwand (Arbeitsentgelt) x_{i2} im Jahr i vermutet:

$$\log Y_i + \theta Y_i = \beta_0 + \beta_1 \log x_{i1} + \beta_2 \log x_{i2} + \varepsilon_i, \quad i = 1, \dots, n$$

Werte in Mio. US-Dollar:

Staat	Produktionsvolumen Y_i	Kapital x_{i1}	Personalaufwand x_{i2}
Ala.	126.148	3.804	31.551
Calif.	3201.486	185.446	452.844
Conn.	690.670	39.712	124.074
Fla.	56.296	6.547	19.181
Ga.	304.531	11.530	45.534
Ill.	723.028	58.987	88.391
Ind.	992.169	112.884	148.530
Iowa	35.796	2.698	8.017
Kans.	494.515	10.360	86.189
Ky.	124.948	5.213	12.000
La.	73.328	3.763	15.900
Maine	29.467	1.967	6.470
Md.	415.262	17.546	69.342
Mass.	241.530	15.347	39.416
Mich.	4079.554	435.105	490.384
Mo.	652.085	32.840	84.831
N.J.	667.113	33.292	83.033
N.Y.	940.430	72.974	190.094
Ohio	1611.899	157.978	259.916
Pa.	617.579	34.324	98.152
Tex.	527.413	22.736	109.728
Va.	174.394	7.173	31.301
Wash.	636.948	30.807	87.963
W. Va.	22.700	1.543	4.063
Wis.	349.711	22.001	52.818

Tabelle 2.4.1

Dabei ist

$$\begin{aligned} g(y, \theta) &= \log y + \theta y & \vec{x} &= (x_1, x_2)^\top \\ f(\vec{x}, \beta) &= \beta_0 + \beta_1 \log x_1 + \beta_2 \log x_2 & \beta &= (\beta_0, \beta_1, \beta_2)^\top \end{aligned}$$

Man sieht, dass $f(x, \beta)$ eine quasilineare Funktion ist, weil sie linear von $\log x_j$ abhängt. $g(y, \theta)$ ist dagegen essentiell nichtlinear.

Wie werden θ und β des Modells (2.4.1) geschätzt? Einerseits ist nach wie vor die MKQ-Methode in Betracht zu ziehen.

$$e(\theta, \beta) = \sum_{i=1}^n (g(Y_i, \theta) - f(\vec{x}_i, \beta))^2 \rightarrow \min_{\theta, \beta}$$

Hier sind $(\hat{\theta}, \hat{\beta}) = \operatorname{argmin} e(\theta, \beta)$ durch numerische nichtlineare Optimierung zu bekommen.

Andererseits bekommt man aber (in der Regel) bessere Schätzer $\hat{\theta}$ und $\hat{\beta}$ für θ und β durch die sog. *Maximum-Likelihood-Methode*.

2.4.1 Maximum-Likelihood-Methode

Es werden folgende Annahmen getroffen:

- die Störgrößen $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$ sind unabhängig voneinander, $i = 1, \dots, n$
- $g(y, \theta)$ ist invertierbar bzgl. y und $g^{-1}(y, \theta)$ stetig differenzierbar mit Ableitung $\neq 0$

Mit Hilfe des Dichte-Transformationsatzes ist dann Y_i auch absolut stetig verteilt mit der Dichte

$$f_{Y_i}(y) = |J| \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp \left\{ -\frac{(g(y, \theta) - f(x, \beta))^2}{2\sigma^2} \right\}, \quad y \in \mathbb{R},$$

wobei

$$X = g(Y_i, \theta) - f(x_i, \beta) = \varepsilon_i \quad \text{und} \quad f_X(u) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp \left\{ -\frac{1}{2} \left(\frac{u}{\sigma} \right)^2 \right\} \quad \text{mit} \quad u = g(y, \theta) - f(x, \beta)$$

und $J = J(y, \theta) = \frac{\partial(g(y, \theta) - f(x, \beta))}{\partial y} = \frac{\partial g(y, \theta)}{\partial y}$ der Jacobian der Transformationsfunktion ist.

Wiederholung des Dichte-Transformationsatzes *WR-Skript: Satz 3.6.1(2):*

Falls X absolut stetig mit Dichte f_X ist und $C \subset \mathbb{R}$ eine offene Menge mit $P(X \in C) = 1$, dann ist $\varphi(X)$ absolut stetig mit Dichte

$$f_{\varphi(X)}(y) = f_X(\varphi^{-1}(y)) \cdot |\varphi^{-1}'(y)|, \quad y \in \varphi(C),$$

falls φ eine auf C stetig differenzierbare Funktion mit $\varphi'(X) \neq 0, x \in C$ ist.

Die Likelihood-Funktion der Stichprobe (Y_1, \dots, Y_n) ist dann gegeben durch

$$L(y_1, \dots, y_n) = \prod_{i=1}^n f_{Y_i}(y_i) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \cdot \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (g(y_i, \theta) - f(\vec{x}_i, \beta))^2\right\} \prod_{i=1}^n \left| \frac{\partial g(y_i, \theta)}{\partial y_i} \right|$$

und die Log-Likelihood-Funktion ist somit

$$\log L(y_1, \dots, y_n) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log \sigma^2 + \sum_{i=1}^n \log |J(y_i, \theta)| - \frac{1}{2\sigma^2} \sum_{i=1}^n (g(y_i, \theta) - f(\vec{x}_i, \beta))^2. \quad (2.4.2)$$

Die Maximum-Likelihood-Schätzer $\hat{\theta}$, $\hat{\beta}$ und $\hat{\sigma}^2$ bekommt man durch

$$\log L(y_1, \dots, y_n) \rightarrow \max_{\theta, \beta, \sigma^2}.$$

Vorausgesetzt, dass dieses Maximum existiert und eindeutig ist, findet man die Lösung durch folgendes nichtlineares Gleichungssystem

$$\begin{aligned} \frac{\partial \log L}{\partial \sigma^2} &= -\frac{n}{2} \cdot \frac{1}{\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_{i=1}^n (g(y_i, \theta) - f(\vec{x}_i, \beta))^2 = 0, \quad i = 1, \dots, n \\ \Rightarrow \hat{\sigma}^2 &= \frac{1}{n} \sum_{i=1}^n (g(y_i, \theta) - f(\vec{x}_i, \beta))^2, \end{aligned}$$

wobei an Stelle von θ und β jeweils ML-Schätzer $\hat{\theta}$ und $\hat{\beta}$ eingesetzt werden sollen, die man aus weiteren Gleichungen bekommt:

$$\begin{aligned} \frac{\partial \log L}{\partial \beta_j} &= -\frac{1}{2\sigma^2} 2(-1) \sum_{i=1}^n (g(y_i, \theta) - f(\vec{x}_i, \beta)) \cdot \frac{\partial f(\vec{x}_i, \beta)}{\partial \beta_j} = 0 \\ &\Leftrightarrow \sum_{i=1}^n (g(y_i, \theta) - f(\vec{x}_i, \beta)) \cdot \frac{\partial f(\vec{x}_i, \beta)}{\partial \beta_j} = 0, \quad j = 1, \dots, m \\ \frac{\partial \log L}{\partial \theta_j} &= \sum_{i=1}^n \frac{1}{|J(y_i, \theta)|} \frac{\partial |J(y_i, \theta)|}{\partial \theta_j} - \frac{1}{2\sigma^2} 2 \sum_{i=1}^n (g(y_i, \theta) - f(\vec{x}_i, \beta)) \frac{g(y_i, \theta)}{\partial \theta_j} = 0, \quad j = 1, \dots, k. \end{aligned}$$

Dieses Gleichungssystem muss numerisch gelöst werden (z.B. mit Hilfe der Statistik-Software R, SAS, S Plus, SPSS).

Da $\hat{\sigma}^2$ bei fixierten θ und β bekannt ist, kann man $\hat{\sigma}^2$ an Stelle von σ^2 in das o.g. Gleichungssystem einsetzen. Man bekommt die sog. *konzentrierte Log-Likelihood-Gleichungen*:

$$\begin{aligned} \log L_c \Big|_{\sigma^2=\hat{\sigma}^2} &= -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log \left(\frac{1}{n} \sum_{i=1}^n (g(y_i, \theta) - f(\vec{x}_i, \beta))^2 \right) + \sum_{i=1}^n \log L |J(y_i, \theta)| - \frac{1}{2(\frac{1}{n})} \\ &= -\frac{n}{2} (\log(2\pi) + 1) - \frac{n}{2} \log \left(\frac{1}{n} \sum_{i=1}^n (g(y_i, \theta) - f(\vec{x}_i, \beta))^2 \right) + \sum_{i=1}^n \log |J(y_i, \theta)| \\ &\rightarrow \max_{\theta, \beta} . \end{aligned} \tag{2.4.3}$$

Beispiel 2.4.1 (Verallgemeinerte Produktionsfunktion - Fortsetzung)

$$\log Y_i + \theta Y_i = \beta_0 + \beta_1 \log x_{i1} + \beta_2 \log x_{i2} + \varepsilon_i, \quad i = 1, \dots, n$$

Zielstellung: Bestimmung von $\hat{\theta}$, $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\sigma}^2$

Falls $\theta = 0$, dann kann das quasilineare Modell (siehe oben) mit neuen Variablen $\log y_i$, $\log x_{ij}$ durch die klassische MKQ-Methode untersucht werden. Vorteile von $\theta \neq 0$?

Berechnen wir die Log-Likelihood-Funktion durch Einsetzen in (2.4.2) :

$$\begin{aligned} \log L(y_1, \dots, y_n) &= -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma^2 + \sum_{i=1}^n \log \left(\frac{1 + \theta y_i}{y_i} \right) \\ &\quad - \frac{1}{2\sigma^2} \sum_{i=1}^n (\log y_i + \theta y_i - \beta_0 - \beta_1 \log x_{i1} - \beta_2 \log x_{i2})^2 \\ &\rightarrow \max_{\sigma^2, \beta, \theta} , \end{aligned}$$

$$\text{weil } J(y_i, \theta) = \frac{\partial g(y_i, \theta)}{\partial y_i} = y_i^{-1} + \theta = \frac{\theta y_i + 1}{y_i}, \quad i = 1, \dots, n .$$

Vorgehensweise beim Optimierungsproblem:

1. Wähle das maximale $\theta_0 : \theta \in [0, \theta_0]$. Im Intervall $[0, \theta_0]$ wähle ein Gitter

$$\theta_j = \frac{j}{M} \cdot \theta_0, \quad j = 0, \dots, M,$$

wobei M vorgegeben ist.

2. Fixiere $\theta = \theta_j$. Bestimme MKQ-Schätzer für β und σ^2 aus dem quasilinearen Modell.

$$\underbrace{\log Y_i + \theta_j Y_i}_{\tilde{Y}_i} = \beta_0 + \beta_1 \underbrace{\log x_{i1}}_{\tilde{x}_{i1}} + \beta_2 \underbrace{\log x_{i2}}_{\tilde{x}_{i2}} + \varepsilon_i, \quad i = 1, \dots, n \quad (n = \text{Anzahl der Staaten})$$

Das Ergebnis der MKQ-Methode sei $\hat{\beta}_j, \hat{\sigma}_j^2, j = 0, \dots, M$.

3. Finde $j_0 = 0, \dots, M : (\theta_{j_0}, \beta_{j_0}, \sigma_{j_0}^2) = \underset{j=0, \dots, M}{\operatorname{argmax}} \log L(y_i, \dots, y_n)$.

4. Bei Bedarf vergrößere M und führe Schritte 1 bis 3 nochmals durch.

5. Setze $(\hat{\theta}, \hat{\beta}_j, \hat{\sigma}^2) = (\theta_{j_0}, \beta_{j_0}, \sigma_{j_0}^2)$.

Lösung:

$$\hat{\theta} = 0.134, \quad \hat{\sigma}^2 = 0.0485, \quad \hat{\beta}_0 = 3.0129, \quad \hat{\beta}_1 = 0.333, \quad \hat{\beta}_2 = 1.1551$$

Was passiert, wenn man die MKQ-Methode direkt bei $\theta = 0$ anwendet?

\Rightarrow Die Parameter β_1 und β_2 , die die Steigung der Regressionsebene in diesem Fall angeben, sind

$$\tilde{\beta}_1 = 0.279, \quad \tilde{\beta}_2 = 0.927 \quad (\text{deutlich andere Steigung als vorher!})$$

Eine noch genauere Analyse wäre durch den mittleren quadratischen Fehler gegeben!

2.4.2 Box-Cox-Transformation

Wie kann man aus linearen Modellen, die keine zufriedenstellende Modellierungsgenauigkeit liefern, auf eine einfache Art und Weise nichtlineare Modelle mit Parametern basteln? \Rightarrow *Box-Cox-Transformation*

Definition 2.4.1

Sei $x > 0$. Die *Box-Cox-Transformation* von x mit Parameter $\lambda \in \mathbb{R}$ ist gegeben durch

$$x^{(\lambda)} = \begin{cases} \frac{x^\lambda - 1}{\lambda}, & \text{wenn } \lambda \neq 0 \\ \log x, & \text{wenn } \lambda = 0 \end{cases}, \quad \text{wobei } x^0 \text{ per Stetigkeit definiert ist, denn}$$

$$\lim_{\lambda \rightarrow 0} x^{(\lambda)} = \lim_{\lambda \rightarrow 0} \frac{x^\lambda - 1}{\lambda} \stackrel{\text{L'Hôpital}}{=} \lim_{\lambda \rightarrow 0} \frac{\log x \cdot x^\lambda}{1} = \log x \quad \text{mit } x^\lambda = e^{\lambda \log x}.$$

Gegeben sei ein lineares Regressionsmodell

$$Y_i = \sum_{j=1}^m x_{ij} \beta_j + \varepsilon_i, \quad i = 1, \dots, n,$$

das den Zusammenhang zwischen Y_i und x_{ij} noch sehr ungenau beschreibt.

Wie kann man dieses Modell verbessern? Dazu führen wir das neue nichtlineare Modell

$$Y_i^{(\theta)} = \sum_{j=1}^m x_{ij}^{(\lambda_j)} \beta_j + \varepsilon_i, \quad i = 1, \dots, n, \quad (2.4.4)$$

ein, wobei $Y_i^{(\theta)}$ bzw. $x_{ij}^{(\lambda_j)}$ die Box-Cox-Transformationen der Ziel- bzw. Ausgangsvariablen Y_i und x_{ij} mit Parametern $\theta, \lambda_1, \dots, \lambda_m$ sind.

⇒ Das neue Modell ist flexibler mit zusätzlichen $m + 1$ Parametern, womit der neue Parametervektor durch

$$(\beta_1, \dots, \beta_m, \theta, \lambda_1, \dots, \lambda_m)$$

gegeben ist. Dieser lässt sich mit Hilfe der bekannten Methoden (MKQ, ML) bestimmen. In der Praxis wählt man oft $\lambda_1 = \dots = \lambda_m = \lambda$. Sehr oft setzt man $\theta = \lambda$.

Spezialfälle:

$$\theta = \lambda = 1 \Rightarrow \text{lineares Modell}$$

$$\theta = \lambda = 0 \Rightarrow \text{log-lineares Modell (quasilinear)}$$

Der Parameterbereich für θ, λ ist meistens $[-2, 2]$. Gesucht wird ein Schätzer $\hat{\theta}, \hat{\lambda}, \hat{\beta}$ für θ, λ, β

Vorgehensweise im Spezialfall $\theta = 1$ (linke Seite ist linear!):

$$Y_i = \sum_{j=1}^m x_{ij}^{(\lambda)} \beta_j + \varepsilon_i, \quad i = 1, \dots, n$$

Für ein konkretes λ ist dies ein quasilineares Regressionsmodell! Darauf basiert folgender Algorithmus:

1. Wähle ein Gitter G der Werte λ aus $[-2;2]$ (z.B. mit Schrittweite $\Delta\lambda = 0.01$).
2. Für jedes λ_0 aus diesem Gitter G finde einen MKQ-Schätzer $\hat{\beta}(\lambda_0)$ für $\beta = (\beta_1, \dots, \beta_m)$.

Sei

$$S(\lambda_0) = \sum_{i=1}^n \left(Y_i - \sum_{j=1}^m x_{ij}^{(\lambda_0)} \hat{\beta}_j(\lambda_0) \right)^2$$

die Summe der quadrierten Residuen.

3. Wähle $\lambda_0 = \underset{\lambda_0 \in G}{\operatorname{argmin}} S(\lambda_0)$. Setze $\hat{\lambda} = \lambda_0$.

4. Der dazugehörige Parametervektor $\hat{\beta}(\lambda_0)$ sei unser Schätzer $\hat{\beta}$.

Beispiel 2.4.2 (Geldbedarf)

vgl. Greene - Econometric Analysis, S. 323ff

Y_i = Geldbedarf (Mio. USD) im Jahr i

x_{i1} = Zinssatz der N.Y. Federal Reserve Bank im Dezember des Jahres i

x_{i2} = Bruttosozialprodukt im Jahr i

Jahr	Geldbedarf Y_i	Zinssatz x_{i1}	Bruttosozialprodukt x_{i2}
1966	480.0	4.50	2208.3
1967	524.3	4.19	2271.4
1968	566.3	5.16	2365.6
1969	589.5	5.87	2423.3
1970	628.2	5.95	2416.2
1971	712.8	4.88	2484.8
1972	805.2	4.50	2608.5
1973	861.0	6.44	2744.1
1974	908.4	7.83	2729.3
1975	1023.1	6.25	2695.0
1976	1163.6	5.50	2826.7
1977	1286.6	5.46	2958.6
1978	1388.9	7.46	3115.2
1979	1497.9	10.28	3192.4
1980	1631.4	11.77	3187.1
1981	1794.4	13.42	3248.8
1982	1954.9	11.02	3166.0
1983	2188.8	8.50	3277.7
1984	2371.7	8.80	3492.0
1985	2563.6	7.69	3573.5

Tabelle 2.4.2a

Lineares Modell:

$$Y_i = \beta_0 + x_{i1}\beta_1 + x_{i2}\beta_2 + \varepsilon_i, \quad i = 1, \dots, n \quad (n = 20 \text{ Jahre: von 1966 bis 1985})$$

⇒ MKQ-Schätzer $(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2) = (-3169.42, -14.9223, 1.58825)$

Log-lineares Modell:

$$\log Y_i = \beta_0 + \log x_{i1}\beta_1 + \log x_{i2}\beta_2 + \varepsilon_i, \quad i = 1, \dots, n$$

⇒ MKQ-Schätzer $(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2) = (-21.992, -0.0315, 3.65628)$

Box-Cox-Transformiertes Modell:

$$\log Y_i = \beta_0 + x_{i1}^{(\lambda)}\beta_1 + x_{i2}^{(\lambda)}\beta_2 + \varepsilon_i, \quad i = 1, \dots, n$$

λ - neuer Parameter

λ_0	0.30	0.40	0.41	0.42	...	0.46	0.47	0.48	0.49
$S(\lambda_0)$	0.13016	0.12732	0.12729	0.12726	...	0.12721	0.12720	0.12721	0.12721

Tabelle 2.4.2b

$$\Rightarrow \hat{\lambda} = 0.47$$

Für $\lambda = \hat{\lambda} = 0.47$ ist der entsprechende MKQ-Schätzer für β gegeben durch

$$\hat{\beta} = (-0.543, -0.00607, -0.0867)$$

Wie schätzt man λ und β im Modell

$$Y_i^{(\lambda)} = \sum_{j=1}^m x_{ij}^{(\lambda)}\beta_j + \varepsilon_i, \quad i = 1, \dots, n$$

Annahme: $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$ unabhängig, $i = 1, \dots, n$

Wenden wir die Maximum-Likelihood-Methode für nichtlineare Modelle mit

$$g(Y_i, \lambda) = Y_i^{(\lambda)}$$

$$f(\vec{x}_i, \beta) = \sum_{j=1}^m \beta_j x_{ij}^{(\lambda)}$$

an. Die Log-Likelihood-Funktion aus (2.4.2) schreibt sich in diesem Fall

$$\log L(y_1, \dots, y_n | \lambda, \beta) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log \sigma^2 + \underbrace{(\lambda - 1) \sum_{i=1}^n \log y_i}_{=\sum_{i=1}^n \log |J(y_i)|} - \frac{1}{2\sigma^2} \underbrace{\sum_{i=1}^n \left(Y_i^{(\lambda)} - \sum_{j=1}^m \beta_j x_{ij}^{(\lambda)} \right)^2}_{=S(\lambda)},$$

weil

$$J(y_i) = \frac{\partial g(y_i, \lambda)}{\partial y_i} = \frac{\partial \left(\frac{y_i^\lambda - 1}{\lambda} \right)}{\partial y_i} = \frac{\lambda y_i^{\lambda-1}}{\lambda} = y_i^{\lambda-1},$$

$$\Rightarrow Y_i > 0 \Rightarrow \log |J(y_i)| = (\lambda - 1) \log y_i, \quad i = 1, \dots, n.$$

Die konzentrierte Log-Likelihood-Funktion ist dann gegeben durch

$$\log L_c(y_1, \dots, y_n | \lambda, \beta) = -\frac{n}{2} (\log(2\pi) + 1) + (\lambda - 1) \sum_{i=1}^n \log Y_i - \frac{n}{2} \log \left(\frac{S(\lambda)}{n} \right) \rightarrow \max_{\lambda, \beta}$$

\Rightarrow Um den ML-Schätzer $(\hat{\lambda}, \hat{\beta}) = \underset{\theta, \beta}{\operatorname{argmax}} \log L_c(y_1, \dots, y_n | \lambda, \beta)$ zu bekommen, geht man wie folgt vor:

1. Wähle ein Gitter G der Parameterwerte λ (siehe 1. Schritt des letzten Algorithmus)
2. Für jedes $\lambda_0 \in G$ finde $\hat{\beta}(\lambda_0) = \underset{\beta}{\operatorname{argmax}} \log L_c(y_1, \dots, y_n | \lambda_0, \beta)$
3. Gebe $\hat{\lambda} = \underset{\lambda_0 \in G}{\operatorname{argmax}} \log L_c(y_1, \dots, y_n | \lambda_0, \hat{\beta}(\lambda_0))$ und $\hat{\beta} = \hat{\beta}(\hat{\lambda})$ als Ergebnis des Algorithmus aus

Beispiel 2.4.2 (Geldbedarf - Fortsetzung)

Wir betrachten das Modell

$$Y_i^{(\lambda)} = \beta_0 + \beta_1 x_{i1}^{(\lambda)} + \beta_2 x_{i2}^{(\lambda)} + \varepsilon_i, \quad i = 1, \dots, n \quad (n = 20)$$

Wir finden die ML-Schätzer $\hat{\lambda}$ und $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$ wie oben beschrieben:

$$\hat{\lambda} = -0.35, \quad \hat{\beta}_0 = -11.170, \quad \hat{\beta}_1 = -0.005689, \quad \hat{\beta}_2 = 5.1437$$

2.5 Test der Linearität

In diesem Abschnitt beschäftigen wir uns mit der Frage, welches Modell besser zu den ökonomischen Daten passt.

2.5.1 Reset-Tests

Wie kann man allgemein die Annahme der Linearität testen?

Beispiel der einfachen (linearen) Regression:

$$Y_i = \alpha + \beta x_i + \varepsilon_i, \quad i = 1, \dots, n \quad (2.5.1)$$

$$Y_i = f(x_i) + \varepsilon_i, \quad i = 1, \dots, n \quad (2.5.2)$$

Es soll das lineare Modell (2.5.1) mit einem nichtlinearen Modell (2.5.2) verglichen werden. Falls das lineare Modell die Daten gut beschreibt, dann gibt es auf der rechten Seite von (2.5.1) keine Terme der höheren Ordnung, d.h. kein x_i^2 , x_i^3 , x_i^4 , usw.

Entwickeln wir $f(x)$ in die Taylor-Reihe um den Punkt $x = 0$:

$$f(x) = f(0) + f'(0)x + \frac{f''(0)}{2!}x^2 + \frac{f'''(0)}{3!}x^3 + \dots$$

Man sieht, dass die rechte Seite (2.5.2) im nichtlinearen Modell von x_i^2 , x_i^3 , x_i^4 , usw. abhängt. Daher kommt die Testregel:

Erweitere die rechte Seite (2.5.1) des linearen Modells um die Ausgangsgrößen

$$x_i^2, x_i^3, x_i^4 : Y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \beta_4 x_i^4 + \varepsilon_i, \quad i = 1, \dots, n$$

und teste, ob $\beta_2 = \beta_3 = \beta_4 = 0$ (im Falle der Linearität).

Diese Vorgehensweise ist im Falle der einfachen Regression sehr elegant und führt zu einem brauchbaren Linearitätstest. Im multivariaten Fall jedoch enthält die Taylor-Entwicklung von

$$f(\vec{x}, \beta), \quad \vec{x} = (x_1, \dots, x_n)^\top$$

auch alle Produkte der Form $x_1^{i_1} \cdot \dots \cdot x_n^{i_n}$. Dadurch wird die Anzahl zusätzlicher Variablen, die in das lineare Modell übernommen werden, sehr hoch, was das Testen der Linearitätsannahme unnötig erschwert.

Betrachten wir diesen Fall z.B. für das folgende lineare Modell

$$Y_i = f(\vec{x}, \beta) + \varepsilon_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i, \quad i = 1, \dots, n.$$

Die Taylor-Reihe zweiter Ordnung der Funktion $f(\vec{x}, \beta)$ hätte in der Umgebung $\vec{x} = \vec{x}^0$ dann folgende Form:

$$f(\vec{x}, \beta) \approx f(\vec{x}, \beta^0) + \frac{\partial f(\vec{x}, \beta)}{\partial x_1}(x_1 - x_1^0) + \frac{\partial f(\vec{x}, \beta)}{\partial x_2}(x_2 - x_2^0) + \frac{1}{2} \left(\frac{\partial^2 f(\vec{x}, \beta)}{\partial^2 x_1}(x_1 - x_1^0)^2 + 2 \cdot \frac{\partial^2 f(\vec{x}, \beta)}{\partial x_1 \partial x_2}(x_1 - x_1^0)(x_2 - x_2^0) + \frac{\partial^2 f(\vec{x}, \beta)}{\partial^2 x_2}(x_2 - x_2^0)^2 \right)$$

Wir bekämen somit ein Modell der Form:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i1}^2 + \beta_4 x_{i1} x_{i2} + \beta_5 x_{i2}^2 + \varepsilon_i, \quad i = 1, \dots, n.$$

Deswegen wird folgende alternative Vorgehensweise angeboten:

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta} x_i, \quad i = 1, \dots, n \quad \text{im linearen Modell}$$

$$\Rightarrow \hat{Y}_i^2 = (\hat{\alpha} + \hat{\beta} x_i)^2 = \hat{\alpha}^2 + 2\hat{\alpha}\hat{\beta} x_i + \hat{\beta}^2 x_i^2 \quad \text{hängt von } x_i^2 \text{ ab.}$$

Genauso hängt \hat{Y}_i^k von x_i^k und den Produkten ab, $k = 2, 3, 4$.

Statt x_i^2 , x_i^3 , x_i^4 einzuführen, ergänze die rechte Seite von (2.5.1) um neue Ausgangsvariablen \hat{Y}_i^2 , \hat{Y}_i^3 , \hat{Y}_i^4 :

$$Y_i = \beta_0 + \beta_1 x_i + \beta_2 \hat{Y}_i^2 + \beta_3 \hat{Y}_i^3 + \beta_4 \hat{Y}_i^4 + \varepsilon_i \quad i = 1, \dots, n.$$

Es soll

$$H_0 : \beta_2 = \beta_3 = \beta_4 = 0 \text{ vs. } H_1 : \exists j \in \{2, 3, 4\} : \beta_j \neq 0$$

getestet werden.

Allgemeine Vorgehensweise:

Es wird getestet, ob der in den Daten vorliegende Zusammenhang zwischen der Zielvariablen Y_i und den Ausgangsvariablen x_{i1}, \dots, x_{im} , $i = 1, \dots, n$, linear ist:

$$Y_i = f(\vec{x}_i, \beta) + \varepsilon_i, \quad i = 1, \dots, n$$

$$H_0 : f(\vec{x}, \beta) = \sum_{j=1}^m x_j \beta_j \text{ (Linearität) vs. } H_1 : f(\vec{x}, \beta) \neq \sum_{j=1}^m x_j \beta_j \text{ (keine Linearität)}$$

1. Sei $\hat{\beta}$ der MKQ-Schätzer von β im linearen Modell

$$Y_i = \sum_{j=1}^m x_{ij}\beta_j + \varepsilon_i, \quad i = 1, \dots, n. \quad \text{Bilde } \hat{Y}_i = \sum_{j=1}^m x_{ij}\hat{\beta}_j, \quad i = 1, \dots, n.$$

2. Ergänze das obige lineare Modell um 3 neue Ausgangsvariablen

$$\hat{Y}_i^2, \hat{Y}_i^3, \hat{Y}_i^4 : Y_i = \sum_{j=1}^{m+3} x_{ij}\beta_j + \varepsilon_i, \quad i = 1, \dots, n,$$

wobei

$$x_{im+1} = \hat{Y}_i^2 \quad x_{im+2} = \hat{Y}_i^3 \quad x_{im+3} = \hat{Y}_i^4$$

3. Berechne die Summe der Residuen in den Fällen 1 und 2 :

$$S = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^n \left(Y_i - \sum_{j=1}^m x_{ij}\hat{\beta}_j \right)^2$$

$$S^* = \sum_{i=1}^n (Y_i - \hat{Y}_i^*)^2 = \sum_{i=1}^n \left(Y_i - \sum_{j=1}^{m+3} x_{ij}\hat{\beta}_j \right)^2$$

Es gilt $S^* \leq S$.

4. Bilde die *Teststatistik*

$$T(Y_1, \dots, Y_n, X) = \frac{(S - S^*)/3}{(S^*)/(n - (m + 3) - 1)} = \frac{S - S^*}{3S^*} (n - m - 4)$$

Es wird getestet:

$$H_0 : \beta_{m+1} = \beta_{m+2} = \beta_{m+3} = 0 \text{ vs. } H_1 : \exists j \in \{m+1, m+2, m+3\} : \beta_j \neq 0$$

Unter H_0 gilt: $T(Y_1, \dots, Y_n, X) \sim F_{3, n-m-4}$

(F-Verteilung: $Z \sim F_{s,t}$ mit $s, t \in \mathbb{N} \Leftrightarrow Z = \frac{\chi_s^2/s}{\chi_t^2/t}$)

5. *Entscheidungsregel*: Lehne H_0 ab, falls

$$T(y_1, \dots, y_n, X) > F_{3, n-m-4, 1-\alpha},$$

wobei das Quantil $F_{3, n-m-4, 1-\alpha}$ der $F_{3, n-m-4}$ -Verteilung zum Niveau $1 - \alpha$ den Tabellen oder einem Statistik-Programm wie z.B. R zu entnehmen ist.

Beispiel 2.1.2 (Milchproduktion - Fortsetzung)

$$Y_i = \alpha + \beta x_i + \varepsilon_i, \quad i = 1, \dots, n \quad (n = 12)$$

Zeigen wir mit Hilfe des Reset-Tests, dass die einfache lineare Regression in diesem Datenbeispiel ein falscher Modellansatz ist. Im erweiterten Modell

$$Y_i = \alpha + \beta_1 x_i + \beta_2 \hat{Y}_i^2 + \beta_3 \hat{Y}_i^3 + \beta_4 \hat{Y}_i^4 \varepsilon_i^*, \quad i = 1, \dots, 12$$

wird

$$H_0 : \beta_2 = \beta_3 = \beta_4 = 0 \text{ (Linearität)} \text{ vs. } H_1 : \exists j \in \{2, 3, 4\} : \beta_j \neq 0 \text{ (keine Linearität)}$$

getestet, wobei

$$\hat{Y}_i = \hat{\alpha} - \hat{\beta} x_i, \quad i = 1, \dots, n$$

und $\hat{\alpha}$, $\hat{\beta}$ MKQ-Schätzer von α und β im ursprünglichen linearen Modell sind.

Summe der quadrierten Residuen im ursprünglichen Modell:

$$S = \sum_{i=1}^{12} (Y_i - \hat{Y}_i)^2$$

Summe der quadrierten Residuen im ergänzten Modell:

$$S^* = \sum_{i=1}^{12} \left(Y_i - \hat{Y}_i - \sum_{j=2}^4 \beta_j \hat{Y}_i^j \right)^2$$

i	1	2	3	4	5	6	7	8	9	10	11	12
\hat{Y}_i	6174	8552	7364	8909	5580	7601	5937	6650	7958	5104	7007	8315

Tabelle 2.5.1

Nach der Berechnung mit Hilfe der Tabelle 2.5.1 bekommt man

$$S = 2786870, \quad S^* = 932014,$$

somit ist die Testgröße gleich

$$T = \frac{(S - S^*)/3}{S^*/(12 - 4 - 1)} = \frac{(2786870 - 932014)/3}{932014/7} \approx 4.644 .$$

Falls das Konfidenzniveau $1 - \alpha = 0.95$ gewählt wird, kann das Quantil der $F_{3,7}$ -Verteilung zum Niveau 0.95 als $F_{3,7,0.95} = 4.347$ aus den Quantiltabellen bestimmt werden.

Da $T = 4.644 > 4.347$, wird die Hypothese H_0 (Linearität) abgelehnt.

2.5.2 Welches Modell ist besser? - Bestimmtheitsmaß R^2

In diesem Abschnitt wird eine Kennzahl R^2 der (quasi)linearen Regressionsmodelle eingeführt, mit deren Hilfe sie verglichen werden können.

Definition 2.5.1

Sei

$$Y_i = \sum_{j=1}^m x_{ij}\beta_j + \varepsilon_i, \quad i = 1, \dots, n$$

ein lineares Modell mit dem MKQ-Schätzer $\hat{\beta}_j$ für β_j , $j = 1, \dots, m$.

Das *Bestimmtheitsmaß* R^2 ist gegeben durch

$$R^2 = \frac{\sum_{i=1}^n (\hat{Y}_i - \bar{Y}_n)^2}{\sum_{i=1}^n (Y_i - \bar{Y}_n)^2},$$

wobei

$$\hat{Y}_i = \sum_{j=1}^m \hat{\beta}_j x_{ij} \quad \text{und} \quad \bar{Y}_n = \frac{1}{n} \sum_{i=1}^n Y_i .$$

R^2 gibt den Anteil der Streuungsreduktion, die durch die Regression entsteht, aus der Gesamtstreuung der Stichprobe (Y_1, \dots, Y_n) an.

Es gilt:

$$R^2 = 0 \Rightarrow \text{keine lineare Abhängigkeit zwischen } Y_i \text{ und } \{x_{ij}\}_{j=1}^m$$

$$R^2 = 1 \Rightarrow \text{perfekter linearer Zusammenhang}$$

Allgemein:

$0 \leq R^2 \leq 1$. Deshalb: je größer der Wert von R^2 ist, desto besser ist unser Modell an die Daten angepasst.

Mehr über R^2 erfährt man in dem Skript „Statistik I“, S. 37-39 oder im Skript „Wirtschaftsstatistik“.

Problemstellung:

Wie vergleiche ich mehrere (quasi)lineare Modelle miteinander mit Hilfe von R^2 ?

⇒ Das Modell mit dem größten Bestimmtheitsmaß R^2 ist das beste, vorausgesetzt, dass die Modelle miteinander verglichen werden dürfen.

Welche Modelle dürfen verglichen werden?

1. Die Modelle sollten dieselbe Zielvariable Y_i besitzen. (Vergleich von unterschiedlichen Maßeinheiten ist nicht gestattet: ein unzulässiger Vergleich wäre z.B. lineares vs. loglineares Modell)
2. Die Modelle sollten dieselbe Anzahl von unabhängigen Variablen besitzen.
3. Die Modelle sollten denselben Niveauparameter α besitzen.

Beispiel 2.1.2 (Milchproduktion - Fortsetzung)

Vergleiche folgende Modelle im Falle der Milchproduktion:

Name	Funktionstyp	R^2
linear	$Y_i = \alpha + \beta x_i (+ \varepsilon_i)$	85.6
! semi-logarithmisch	$Y_i = \alpha + \beta \log x_i (+ \varepsilon_i)$	90.5
invers	$Y_i = \alpha + \frac{\beta}{x_i} (+ \varepsilon_i)$	59.5
exponential	$\log Y_i = \alpha + \beta x_i (+ \varepsilon_i)$	82.0
! logarithmisch	$\log Y_i = \alpha + \beta \log x_i (+ \varepsilon_i)$	94.0
logarithmisch-invers	$\log Y_i = \alpha + \frac{\beta}{x_i} (+ \varepsilon_i)$	67.6
! quadratisch	$Y_i = \alpha + \beta_2 x_i + \beta_3 x_i^2 (+ \varepsilon_i)$	95.0

Tabelle 2.5.2

Fazit: Das semi-logarithmische, logarithmische und quadratische Modell kommen nach unserem Vergleich in Frage.

2.5.3 Box-Cox-Verfahren

Wie kann man Modelle unterschiedlicher Typen miteinander vergleichen, z.B. das lineare gegen das logarithmische Modell (Zielvariablen auf unterschiedlichen Skalen)?

Seien $Y_i > 0$, $i = 1, \dots, n$.

1. Skalierung: Berechne das geometrische Mittel der Stichprobe (Y_1, \dots, Y_n)

$$\tilde{Y}_n = \sqrt[n]{\prod_{i=1}^n Y_i} \quad .$$

Bilde die neuen Zielvariablen $Y_i^* = Y_i / \tilde{Y}_n$, $i = 1, \dots, n$.

2. Um ein (quasi)lineares Modell mit den Zielvariablen Y_i gegen ein weiteres (quasi)lineares Modell mit den Zielvariablen $\log Y_i$ zu vergleichen, setze Y_i^* statt Y_i in diese Modelle ein und führe die MKQ-Schätzung vom Parametervektor β durch.

3. Bilde S_0 und S_1 - die Summe der quadrierten Residuen in den beiden Modellen.

4. - Falls $S_0 < S_1$, dann ist das Modell mit Zielvariable Y_i besser

- Falls $S_0 = S_1$, dann ist das Modell mit Zielvariable Y_i gleichwertig

- Falls $S_0 > S_1$, dann ist das Modell mit Zielvariable Y_i schlechter

als das Modell mit Zielvariable $\log Y_i$.

5. *Statistischer Test*: Bilde die Teststatistik

$$T = \frac{n}{2} \left| \log \frac{S_0}{S_1} \right|$$

für die Hypothesen

H_0 : Modelle gleichwertig vs. H_1 : Modelle nicht gleichwertig .

Unter der Hypothese H_0 gilt $T \sim \chi_1^2$. Somit wird H_0 verworfen, falls

$$T > \chi_{1,1-\alpha}^2 ,$$

wobei $\chi_{1,1-\alpha}^2$ das $(1 - \alpha)$ -Quantil der χ_1^2 -Verteilung ist.



G. E. P. Box
(1919 -)



D. R. Cox
(1927 -)

Beispiel 2.1.2 (Milchproduktion - Fortsetzung)

Vergleichen wir das logarithmische und das semi-logarithmische Modell miteinander:

$$\tilde{Y}_n = 6365.29 \Rightarrow Y_i^* = \frac{Y_i}{\tilde{Y}_n}$$

Die neuen Modelle sind somit

$$Y_i^* = \alpha + \beta \log x_i + \varepsilon_i \quad \text{und} \quad \log(Y_i^*) = \alpha + \beta \log x_i + \varepsilon_i$$

$S_0 = 0.03807$ im semi-logarithmischen Modell , $S_1 = 0.02873$ im logarithmischen Modell

$S_1 < S_0 \Rightarrow$ das logarithmische Modell ist besser als das semi-logarithmische Modell.

Box-Cox-Test:

$$T = \frac{12}{2} \left| \log \frac{0.03807}{0.02873} \right| \approx 1.68877 < \chi_{1,0.95}^2 = 3.84146$$

\Rightarrow die Hypothese der Gleichwertigkeit beider Modelle wird nicht abgelehnt.

Bisher konnten wir entweder einen Test der (Log)Linearität durchführen oder (im Falle des Box-Cox-Tests) prüfen, ob ein Modell mit endogenen Variablen Y_i oder $\log Y_i$ besser zu den Daten passt. Mit Hilfe der Box-Cox-Transformation können wir Modelle in einer größeren Klasse vergleichen. Und zwar betrachten wir folgende Modellverbesserung (vgl. Gleichung (2.4.4)):

$$Y_i^{(\theta)} = \sum_{j=1}^m \beta_j x_{ij}^{(\lambda)} + \varepsilon_i, \quad i = 1, \dots, n .$$

Sie enthält für

$$\left. \begin{array}{ll} \theta = \lambda = 1 & \text{das lineare} \\ \theta = 1, \lambda = 0 & \text{das semi-logarithmische} \\ \theta = 0, \lambda = 1 & \text{das Exponential-} \\ \theta = 0, \lambda = -1 & \text{das log-inverse} \\ \theta = 1, \lambda = -1 & \text{das inverse} \end{array} \right\} \text{Modell als Spezialfall.}$$

Natürlich sind auch andere Werte für θ und λ denkbar.

Das **Box-Cox-Verfahren** (*Box, Cox, Zaremska, 1960er Jahre*) besteht aus drei Schritten:

1. Skalieren die Variablen Y_i durch das geometrische Mittel \tilde{Y}_n von Y_1, \dots, Y_n :

$$\tilde{Y}_n = \sqrt[n]{\prod_{i=1}^n Y_i} = e^{\frac{1}{n} \sum_{i=1}^n \log Y_i} .$$

zu $Y_i^* = Y_i / \tilde{Y}_n$ und betrachte das skalierte Modell

$$Y_i^{*(\theta)} = \sum_{j=1}^m \beta_j x_{ij}^{(\lambda)} + \varepsilon_i , \quad i = 1, \dots, n \quad (2.5.3)$$

2. Berechne für alle θ und λ aus einem Gitter auf $[-2;2]$ (z.B. mit Schrittweite 0.1) den MKQ-Schätzer im quasilinearen Modell (2.5.3). Auf Basis dieser Schätzung errechne die Summe der Residuenquadrate

$$S_{\theta, \lambda}^* = \sum_{i=1}^n (Y_i^{*(\theta)} - \sum_{j=1}^m \hat{\beta}_j x_{ij}^{(\lambda)})^2 .$$

3. Das Paar (θ, λ) mit minimaler Summe $S_{\theta, \lambda}^*$ liefert das beste Modell des Box-Cox-Typs für die vorliegenden Daten Y_1, \dots, Y_n und x_{ij} , $i = 1, \dots, n$, $j = 1, \dots, m$.

2.5.4 Likelihood-Ratio-Test

Wie kann die Annahme der (Log)Linearität getestet werden?

Es ist folgendes allgemeines ökonometrisches Modell gegeben:

$$Y_i^{(\theta)} = \sum_{j=1}^m \beta_j x_{ij}^{(\lambda)} + \varepsilon_i , \quad i = 1, \dots, n ,$$

wobei $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$.

Dabei ist das Ziel zu testen, ob das lineare Modell ($\lambda = 1$) bzw. das loglineare Modell ($\lambda = 0$) verträglich mit den Daten ist.

Gegeben $\theta = \lambda$, wollen wir folgende statistische Hypothese testen:

$$H_0 : \lambda = 1 \text{ (lineares Modell) vs. } H_1 : \lambda \neq 1 \text{ (kein lineares Modell)}$$

oder

$$H_0 : \lambda = 0 \text{ (loglineares Modell) vs. } H_1 : \lambda \neq 0 \text{ (kein loglineares Modell)}$$

Schreiben wir beide Hypothesen in einheitliche Bezeichnungen zusammen:

$$H_0 : \lambda = \lambda_0 \text{ vs. } H_1 : \lambda \neq \lambda_0 , \text{ wobei } \lambda_0 \in \{0, 1\}$$

Wir definieren die Teststatistik:

$$T(Y_1, \dots, Y_n, X) = -2(\log L(Y_1, \dots, Y_n | \lambda_0, \hat{\beta}) - \log L(Y_1, \dots, Y_n | \hat{\lambda}, \hat{\beta})) \geq 0 ,$$

wobei $\hat{\lambda}$ und $\hat{\beta}$ die ML-Schätzer für λ und β sind und $X = (x_{i,j}^{(\lambda)})_{i,j=1,\dots,n}$.

Es gilt

$$T(Y_1, \dots, Y_n, X) \geq 0 , \quad \text{weil} \quad L(Y_1, \dots, Y_n | \hat{\lambda}) = \max_{\lambda \in [-2;2]} L(Y_1, \dots, Y_n | \lambda) .$$

Der Name *Likelihood-Ratio-Test* kommt von der Darstellung

$$T(Y_1, \dots, Y_n, X) = -2 \log \frac{L(Y_1, \dots, Y_n | \lambda_0, \tilde{\beta})}{L(Y_1, \dots, Y_n | \hat{\lambda}, \hat{\beta})} .$$

Man kann zeigen, dass unter H_0 gilt : $T(Y_1, \dots, Y_n, X) \sim \chi_1^2 = z^2$, $z \sim \mathcal{N}(0, 1)$

Die Hypothese H_0 wird abgelehnt, falls

$$T(Y_1, \dots, Y_n, X) > \chi_{1,1-\alpha}^2 ,$$

wobei $\chi_{1,1-\alpha}^2$ das $(1 - \alpha)$ -Quantil der χ_1^2 -Verteilung für ein Konfidenzniveau α ist und α die Wahrscheinlichkeit des Fehlers 1. Art = $P(H_0 \text{ ablehnen} | H_0 \text{ richtig})$ ist (z.B. $\alpha = 0.01, 0.05, \dots$).

Bemerkung 2.5.1

Auf diese Weise ist es nicht möglich, das lineare gegen das logarithmische Modell zu testen, denn die „Testgröße“

$$T(Y_1, \dots, Y_n, X) = -2 \log \frac{L(Y_1, \dots, Y_n | \lambda_0, \tilde{\beta})}{L(Y_1, \dots, Y_n | \hat{\lambda}, \hat{\beta})} \neq \chi_1^2$$

ist nicht χ_1^2 -verteilt - sie ist nicht mehr nichtnegativ.

Beispiel 2.4.2 (Geldbedarf - Fortsetzung)

$$Y_i^{(\lambda)} = \beta_0 + \beta_1 x_{i1}^{(\lambda)} + \beta_2 x_{i2}^{(\lambda)} + \varepsilon_i , \quad i = 1, \dots, n \quad (n = 20) , \quad \varepsilon_i \sim \mathcal{N}(0, \sigma^2)$$

Problem:

Ist das (log)lineare Modell hier angemessen oder nicht?

Wir verwenden den Likelihood-Ratio-Test (s.o.):

Nach der Berechnung gilt:

$$\log L(y_1, \dots, y_n, X|\hat{\lambda}) = -116.51 \quad (\hat{\lambda} = -0.35)$$

$$\log L(y_1, \dots, y_n, X|\lambda = 0) = -118.073$$

$$\log L(y_1, \dots, y_n, X|\lambda = 1) = -130.1333$$

$$T_{\lambda=0}(y_1, \dots, y_n, X) = -2(-118.073 + 116.51) = 3.13$$

$$T_{\lambda=1}(y_1, \dots, y_n, X) = -2(-130.133 + 116.51) = 27.25$$

Wähle das Signifikanzniveau $\alpha = 0.05$. Es gilt $\chi_{1,0.95}^2 = 3.84$.

Da $3.13 < 3.84$, $27.25 > 3.84$, wird die Hypothese der Linearität verworfen. Dagegen wird die Loglinearität nicht abgelehnt.

Fazit:

Für die ökonomische Datenanalyse ist in diesem Fall das lineare Modell nicht geeignet. Das loglineare Modell scheint eine brauchbare (alternative) Annahme zu sein.

2.6 Verallgemeinerte lineare Modelle

Eine andere Klasse von ökonomischen Modellen erlaubt einerseits einen beliebigen funktionellen Zusammenhang g zwischen dem Mittelwert der Zielvariablen $\mathbb{E}Y_i$ und dem linearen Teil $X\beta$, der aus linearen Kombinationen der Einträge der Designmatrix $X = (x_{ij})$ und des Parametervektors $\beta = (\beta_1, \dots, \beta_m)^\top$ besteht; andererseits lässt sie andere Verteilungen von Y_i zu, die nicht notwendigerweise auf der Normalverteilung (und Funktionen davon) basieren. So ist es möglich, Daten Y_i zu betrachten, die eine endliche Anzahl von Ausprägungen haben (z.B. „Ja“ und „Nein“ in ökonomischen Meinungsumfragen). Die Klasse aller möglichen Verteilungen wird durch die sog. *Exponentialfamilie* begrenzt, die wir in Kürze einführen werden.

Sei Y_1, \dots, Y_n eine Zufallsstichprobe der Zielvariablen des Modells und sei $X = (x_{ij})_{\substack{i=1, \dots, n \\ j=1, \dots, m}}$ die Designmatrix der Ausgangsvariablen, die hier nicht zufällig sind.

Definition 2.6.1

Das *verallgemeinerte lineare Modell* ist gegeben durch

$$(g(\mathbb{E}Y_1), \dots, g(\mathbb{E}Y_n))^T = X\beta \quad \text{mit } \beta = (\beta_1, \dots, \beta_m)^T, \quad (2.6.1)$$

wobei $g : G \subset \mathbb{R} \rightarrow \mathbb{R}$ die sog. *Linkfunktion* mit dem Definitionsbereich G und $\text{rg}(X) = m$ ist.

Unter der Annahme, dass g explizit bekannt ist, soll hier der Parametervektor β aus (Y_1, \dots, Y_n) geschätzt werden. Wir setzen voraus, dass $Y_i, i = 1, \dots, n$, unabhängig, aber nicht unbedingt identisch verteilt sind. Ihre Verteilung gehört jedoch zur folgenden Klasse von Verteilungen:

Definition 2.6.2

Die Verteilung einer Zufallsvariable Y gehört zur *Exponentialfamilie*, falls es Funktionen $a : \mathbb{R} \times \mathbb{R}_+ \rightarrow \mathbb{R}$ und $b : \Theta \rightarrow \mathbb{R}$ gibt, für die

- im *absolutstetigen Fall* die Dichte von Y gegeben ist durch

$$f_\theta(y) = \exp \left\{ \frac{1}{\tau^2} (y\theta + a(y, \tau) - b(\theta)) \right\}, \quad y \in \mathbb{R} \quad (2.6.2)$$

- im *diskreten Fall* die Zähldichte von Y gegeben ist durch

$$P_\theta(Y = y) = \exp \left\{ \frac{1}{\tau^2} (y\theta + a(y, \tau) - b(\theta)) \right\}, \quad y \in C, \quad (2.6.3)$$

wobei C der (höchstens) abzählbare Wertebereich von Y , τ^2 der sog. *Störparameter*, $\theta \in \Theta \subset \mathbb{R}$ ein Parameter und

$$\Theta = \left\{ \theta \in \mathbb{R} : \int_{\mathbb{R}} \exp \left\{ \frac{y\theta + a(y, \tau)}{\tau^2} \right\} dy < \infty \right\}$$

bzw.

$$\Theta = \left\{ \theta \in \mathbb{R} : \sum_{y \in C} \exp \left\{ \frac{y\theta + a(y, \tau)}{\tau^2} \right\} < \infty \right\}$$

der natürliche Parameterraum ist, der mindestens zwei verschiedene Elemente enthält.

Man kann zeigen, dass Θ ein Intervall ist.

Beispiel 2.6.1

Welche Verteilungen gehören zur Exponentialfamilie?

1. **Normalverteilung:** Falls $Y \sim \mathcal{N}(\mu, \sigma^2)$, dann ist der Erwartungswert μ der uns inter-

essierende Parameter, σ^2 ist dagegen der Störparameter. Es gilt:

$$\begin{aligned} f_{\mu}(y) &= \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(y-\mu)^2}{2\sigma^2}} \\ &= \exp \left\{ -\frac{1}{2} \log(2\pi\sigma^2) - \frac{1}{2} \left(\frac{y^2}{\sigma^2} - \frac{2y\mu}{\sigma^2} + \frac{\mu^2}{\sigma^2} \right) \right\} \\ &= \exp \left\{ \frac{1}{\sigma^2} \left(y\mu - \frac{y^2}{2} - \left(\frac{\mu^2}{2} + \frac{\sigma^2}{2} \log(2\pi\sigma^2) \right) \right) \right\}, \end{aligned}$$

so dass

$$\theta = \mu, \quad \tau = \sigma, \quad a(y, \tau) = -\frac{y^2}{2} - \frac{\sigma^2}{2} \log(2\pi\sigma^2) \quad \text{und} \quad b(\mu) = b(\theta) = \frac{\mu^2}{2}.$$

2. **Bernoulli-Verteilung:** $Y \sim \text{Bernoulli}(p)$, $p \in [0; 1]$.

Sie wird etwa im Falle von Meinungsumfragen in der Marktforschung verwendet, in denen

$$Y = \begin{cases} 1, & \text{falls die Antwort „ja“} \\ 0, & \text{falls die Antwort „nein“} \end{cases} \quad \text{auf eine Frage der Enquete gegeben wurde.}$$

Dabei ist die Wahrscheinlichkeit $P(Y = 1) = p$, $P(Y = 0) = 1 - p$. Dann gilt für $y \in \{0, 1\}$:

$$\begin{aligned} P_{\theta}(Y = y) &= p^y(1-p)^{1-y} = e^{y \log p + (1-y) \log(1-p)} \\ &= e^{y \log \frac{p}{1-p} - (-\log(1-p))}. \end{aligned}$$

Somit gehört die Bernoulli-Verteilung zur Exponentialfamilie mit

$$\theta = \log \frac{p}{1-p}, \quad \tau = 1, \quad a(y, \tau) = 0, \quad b(\theta) = -\log(1-p) = \log(1 + e^{\theta}).$$

3. **Poisson-Verteilung:** Falls $Y \sim \text{Poisson}(\lambda)$, $\lambda > 0$, dann gilt für $y \in \mathbb{N}_0$

$$P_{\theta}(Y = y) = e^{-\lambda} \cdot \frac{\lambda^y}{y!} = e^{y \log \lambda - \log(y!) - \lambda}.$$

Somit gehört die Poisson-Verteilung zur Exponentialfamilie mit

$$\theta = \log \lambda, \quad \tau = 1, \quad a(y, \tau) = -\log(y!), \quad b(\theta) = \lambda = e^{\theta}.$$

Lemma 2.6.1

Falls die Verteilung von Y zur Exponentialfamilie gehört, $\mathbb{E}Y^2 < \infty$ und $b : \Theta \rightarrow \mathbb{R}$ zweimal stetig differenzierbar ist mit $b''(\theta) > 0$ für alle $\theta \in \Theta$, dann gilt

$$\mathbb{E}Y = b'(\theta), \quad \text{Var} Y = \tau^2 b''(\theta).$$

Die Zielgrößen Y_i , $i = 1, \dots, n$ seien also unabhängig verteilt mit einer Verteilung, die zur Exponentialfamilie gehört und einer (Zähl)Dichte wie in (2.6.2) bzw. (2.6.3). Setzen wir voraus, dass $b : \Theta \rightarrow \mathbb{R}$ zweimal stetig differenzierbar ist mit $b''(\theta) > 0$ für alle $\theta \in \Theta$. Sei ein verallgemeinertes lineares Modell (2.6.1) gegeben.

Definition 2.6.3 (Natürliche Linkfunktion)

Die Linkfunktion $g : G \rightarrow \mathbb{R}$ heißt *natürlich*, falls $g = (b')^{-1}$, $G = \{b'(\theta) : \theta \in \Theta\}$ und g zweimal stetig differenzierbar ist mit $g'(x) \neq 0$ für alle $x \in G$. Die Frage, warum die natürliche Linkfunktion so heißt, beantwortet folgendes Lemma:

Lemma 2.6.2

Falls das verallgemeinerte lineare Modell (2.6.1) die natürliche Linkfunktion besitzt, dann gilt $(\theta_1, \dots, \theta_n)^\top = X\beta$.

Beweis

Wegen $b''(\theta) > 0$ ist $b'(\theta)$ monoton steigend, also invertierbar. Führen wir folgende Bezeichnungen ein:

$$\mu_i = \mathbb{E} Y_i, \quad \eta_i = x_i^\top \beta, \quad x_i = (x_{i1}, \dots, x_{im})^\top, \quad i = 1, \dots, n$$

Da g invertierbar ist, gilt

$$\mu_i = g^{-1}(x_i^\top \beta) = g^{-1}(\eta_i), \quad i = 1, \dots, n$$

Andererseits folgt $\mu_i = b'(\theta_i)$ aus Lemma 2.6.1, so dass

$$b'(\theta_i) = g^{-1}(\eta_i) \stackrel{\text{Definition 2.6.3}}{=} b'(\eta_i), \quad i = 1, \dots, n.$$

Wegen der Monotonie von b' folgt die Behauptung $\theta_i = \eta_i$, $i = 1, \dots, n$.

□

Beispiel 2.6.2

Berechnen wir die natürlichen Linkfunktionen für die Verteilungen von Beispiel 2.6.1.

1. **Normalverteilung:** da $b(\mu) = \frac{\mu^2}{2} + \frac{\sigma^2}{2} \log(2\pi\sigma^2)$, gilt

$$b'(x) = \frac{2x}{2} = x \text{ und somit } g(x) = (b')^{-1}(x) = x.$$

Die natürliche Linkfunktion ist $g(x) = x$, somit gilt hier

$$(\mu_1, \dots, \mu_n)^\top = (\mathbb{E} Y_1, \dots, \mathbb{E} Y_n)^\top = X\beta .$$

Das ist genau der Fall der linearen Regression.

2. **Bernoulli-Verteilung:** da $b(\theta) = \log(1 + e^\theta)$, gilt

$$\begin{aligned} b'(x) &= \frac{1}{1 + e^x} \cdot e^x = y \\ \Leftrightarrow \frac{1}{e^{-x} + 1} &= y \\ \Leftrightarrow \frac{1}{y} - 1 &= e^{-x} \\ \Leftrightarrow x &= -\log \frac{1-y}{y} = \log \frac{y}{1-y} \\ \Rightarrow g(x) &= (b')^{-1}(x) = \log \frac{x}{1-x} . \end{aligned}$$

Das verallgemeinerte lineare Regressionsmodell im Falle der Bernoulli-Verteilung wird *binäre (kategoriale) Regression* genannt. Falls sie mit der natürlichen Linkfunktion verwendet wird, nennt man sie *logistische Regression*. In diesem Fall gilt

$$\begin{aligned} (p_1, \dots, p_n)^\top &= (\mathbb{E} Y_1, \dots, \mathbb{E} Y_n)^\top \\ \theta_i &= \log \frac{p_i}{1-p_i} = x_i^\top \beta , \quad i = 1, \dots, n \\ \Leftrightarrow e^{\theta_i} &= \frac{p_i}{1-p_i} \\ \Leftrightarrow p_i &= \frac{e^{\theta_i}}{1 + e^{\theta_i}} \\ \Leftrightarrow p_i &= \frac{e^{x_i^\top \beta}}{1 + e^{x_i^\top \beta}} , \quad i = 1, \dots, n . \end{aligned}$$

Das Verhältnis

$$\frac{p_i}{1-p_i} = \frac{P(Y_i = 1)}{P(Y_i = 0)} , \quad i = 1, \dots, n$$

wird in der englischsprachigen Literatur *Odds* genannt. Der Logarithmus des Odds heißt *Logit*:

$$\log \frac{p_i}{1-p_i} , \quad i = 1, \dots, n .$$

Logits sind also hier „neue Zielvariablen“, die durch Linearkombinationen $x_i^\top \beta$ geschätzt werden.

Eine alternative Linkfunktion, die oft benutzt wird, ist $g(x) = \Phi^{-1}(x)$, die *Quantilfunktion der Normalverteilung*. Sie ist keine natürliche Funktion. Mit ihrer Hilfe bekommt man das sog. *Probit-Modell*:

$$p_i = \Phi(x_i^\top \beta) , \quad i = 1, \dots, n .$$

3. **Poisson-Verteilung:** da $b(\theta) = e^\theta$, ist in diesem Fall

$$g(x) = (b')^{-1}(x) = \log x, \quad x > 0$$

die natürliche Linkfunktion. Somit hat das verallgemeinerte lineare Modell mit der natürlichen Linkfunktion folgende Darstellung

$$(\log \lambda_1, \dots, \log \lambda_n)^\top = X\beta \quad \text{oder} \quad \lambda_i = e^{x_i^\top \beta}, \quad i = 1, \dots, n.$$

2.6.1 Maximum-Likelihood-Schätzung von β

Da die (Zähl)Dichte von Y_i die Gestalt

$$\exp \left\{ \frac{1}{\tau^2} (y\theta_i + a(y, \tau) - b(\theta_i)) \right\}$$

hat und Y_i unabhängig sind, kann man die Log-Likelihood-Funktion der Stichprobe $Y = (Y_1, \dots, Y_n)$ in folgender Form aufschreiben:

$$\log L(Y, \theta) = \log \prod_{i=1}^n f_{\theta_i}(Y_i) = \frac{1}{\tau^2} \sum_{i=1}^n (Y_i \theta_i + a(Y_i, \tau) - b(\theta_i)). \quad (2.6.4)$$

Aus dem Beweis des Lemmas 2.6.2 folgt, dass

$$\theta_i = (b')^{-1}(g^{-1}(x_i^\top \beta)), \quad i = 1, \dots, n,$$

was bedeutet, dass die Funktion $\log L(Y, \theta)$ eine Funktion von Parameter β ist. In der Zukunft schreiben wir $\log L(Y, \beta)$, um diese Tatsache zu unterstreichen.

Unser Ziel ist es, den Maximum-Likelihood-Schätzer $\hat{\beta}$ für β zu berechnen:

$$\hat{\beta} = \underset{\beta}{\operatorname{argmax}} \log L(Y, \beta).$$

Dafür wird die notwendige Bedingung des Extremums

$$\frac{\partial \log L(Y, \beta)}{\partial \beta_i} = 0, \quad i = 1, \dots, m,$$

untersucht. Verwenden wir folgende Bezeichnungen:

$$U_i(\beta) = \frac{\partial \log L(Y, \beta)}{\partial \beta_i}, \quad i = 1, \dots, m,$$

$$U(\beta) = (U_1(\beta), \dots, U_m(\beta))^\top,$$

$$I_{ij}(\beta) = \mathbb{E}[U_i(\beta)U_j(\beta)], \quad i, j = 1, \dots, m.$$

Die Matrix $I(\beta) = (I_{ij}(\beta))_{i,j=1}^m$ heißt *Fisher-Informationsmatrix*. Man kann zeigen, dass $U(\beta)$ und $I(\beta)$ folgende explizite Form haben:

Satz 2.6.1

Es gilt

$$U_j(\beta) = \sum_{i=1}^n x_{ij} (Y_i - \mu_i(\beta)) \frac{\partial g^{-1}(\eta_i)}{\partial \eta_i} \frac{1}{\sigma_i^2(\beta)}, \quad j = 1, \dots, m,$$

$$I_{jk}(\beta) = \sum_{i=1}^n x_{ij} x_{ik} \left(\frac{\partial g^{-1}(\eta_i)}{\partial \eta_i} \right)^2 \frac{1}{\sigma_i^2(\beta)}, \quad j, k = 1, \dots, m, \text{ wobei}$$

$$\eta_i = x_i^\top \beta,$$

$\mu_i(\beta) = g^{-1}(x_i^\top \beta)$ der Erwartungswert von Y_i und

$$\sigma_i^2(\beta) \stackrel{\text{Lemma 2.6.1}}{=} \tau^2 b''(\theta_i) \stackrel{(2.6.5)}{=} \tau^2 b''((b')^{-1}(g^{-1}(x_i^\top \beta))), \quad i = 1, \dots, n \text{ die Varianz von } Y_i \text{ ist.}$$

Ein Beweis befindet sich im „Statistik II“ - Skript, S.90f.

Bemerkung 2.6.1

Im Falle der natürlichen Linkfunktion vereinfachen sich die obigen Gleichungen. So sieht die Log-Likelihood-Funktion folgendermaßen aus:

$$\log L(Y, \beta) = \frac{1}{\tau^2} \sum_{i=1}^n \left(Y_i x_i^\top \beta + a(Y_i, \tau) - b(x_i^\top \beta) \right)$$

Da in diesem Fall $g^{-1}(\eta_i) = b'(\eta_i)$, $\eta_i = x_i^\top \beta = \theta_i$ gilt

$$\frac{\partial g^{-1}(\eta_i)}{\partial \eta_i} = b''(\theta_i) \stackrel{\text{Lemma 2.6.1}}{=} \frac{1}{\tau^2} \sigma_i^2(\beta)$$

und somit

$$U_j(\beta) = \frac{1}{\tau^2} \sum_{i=1}^n x_{ij} (Y_i - \mu_i(\beta)) , \quad j = 1, \dots, m ,$$

$$I_{jk}(\beta) = \frac{1}{\tau^4} \sum_{i=1}^n x_{ij} x_{ik} \sigma_i^2(\beta) , \quad j, k = 1, \dots, m .$$

Definition 2.6.4

Führen wir die sog. *Hesse-Matrix* $W(\beta)$ als zufällige Matrix

$$W(\beta) = (W_{ij}(\beta))_{i,j=1}^m \quad \text{mit} \quad W_{ij}(\beta) = \frac{\partial^2}{\partial \beta_i \partial \beta_j} \log L(Y, \beta)$$

ein. Diese $(m \times m)$ -Matrix enthält die partiellen Ableitungen 2. Ordnung der Log-Likelihood-Funktion, die für die numerische Lösung der Maximierungsaufgabe

$$\log L(Y, \beta) \rightarrow \max_{\beta}$$

von Bedeutung sein werden.

Satz 2.6.2

Es gilt

$$W_{jk}(\beta) = \sum_{i=1}^n x_{ij} x_{ik} \left((Y_i - \mu_i(\beta)) \nu_i - u_i^2 \frac{1}{\sigma_i^2(\beta)} \right) , \quad j, k = 1, \dots, m ,$$

wobei

$$u_i = \frac{\partial g^{-1}(\eta_i)}{\partial \eta_i} \quad \text{und} \quad \nu_i = \frac{1}{\tau^2} \cdot \frac{\partial^2 ((b')^{-1} \circ g^{-1}(\eta_i))}{\partial \eta_i^2} , \quad i = 1, \dots, n .$$

Beweis

Für beliebige $j, k = 1, \dots, m$ gilt

$$\begin{aligned}
W_{jk}(\beta) &= \frac{\partial}{\partial \beta_k} U_j(\beta) = \frac{\partial}{\partial \beta_k} \sum_{i=1}^n x_{ij} (Y_i - \mu_i(\beta)) \frac{\partial g^{-1}(\eta_i)}{\partial \eta_i} \frac{1}{\sigma_i^2(\beta)} \\
&= \sum_{i=1}^n x_{ij} \left((Y_i - \mu_i(\beta)) \frac{\partial}{\partial \beta_k} \left(\frac{\partial g^{-1}(\eta_i)}{\partial \eta_i} \frac{1}{\sigma_i^2(\beta)} \right) - \frac{\partial g^{-1}(\eta_i)}{\partial \eta_i} \frac{1}{\sigma_i^2(\beta)} \frac{\partial \mu_i(\beta)}{\partial \beta_k} \right) \\
&= \sum_{i=1}^n \left(x_{ij} (Y_i - \mu_i(\beta)) \frac{\partial}{\partial \beta_k} \left(\frac{\tau^2 b''((b')^{-1}(g^{-1}(\eta_i))) ((b')^{-1} \circ g^{-1})'(\eta_i)}{\tau^2 b''((b')^{-1}(g^{-1}(\eta_i)))} \right) \right. \\
&\quad \left. - \left(\frac{\partial g^{-1}(\eta_i)}{\partial \eta_i} \right)^2 \frac{1}{\sigma_i^2(\beta)} \right) \\
&= \sum_{i=1}^n x_{ij} x_{ik} \left((Y_i - \mu_i(\beta)) \nu_i - u_i^2 \frac{1}{\sigma_i^2(\beta)} \right) .
\end{aligned}$$

□

Für verallgemeinerte lineare Modelle mit natürlichen Linkfunktionen gilt insbesondere

$$W(\beta) = -I(\beta) = -\frac{1}{\tau^4} \sum_{i=1}^n x_{ij} x_{ik} \sigma_i^2(\beta) , \quad (2.6.5)$$

weil in diesem Fall $\nu_i = 0$ für alle $i = 1, \dots, n$. $W(\beta)$ ist also deterministisch.

Beispiel 2.6.3

Wie sehen $U(\beta)$, $I(\beta)$ und $W(\beta)$ für unsere Modelle aus Beispiel 2.6.2 (natürliche Linkfunktionen) aus?

1. **Normalverteilung:** dieser Fall entspricht der üblichen multivariaten linearen Regression mit normalverteilten Störgrößen. In diesem Fall gilt $\mu = X\beta$, $\tau^2 = \sigma^2$.

Aus Bemerkung 2.6.1 folgt

$$\begin{aligned}
U(\beta) &= \frac{1}{\sigma^2} X^\top (Y - X\beta) , \\
I(\beta) &= \left(\mathbb{E} \left(\frac{\partial \log L(Y, \beta)}{\partial \beta_i} \cdot \frac{\partial \log L(Y, \beta)}{\partial \beta_j} \right) \right)_{i,j=1,\dots,m} = \frac{1}{\sigma^2} X^\top X , \\
W(\beta) &= -I(\beta) .
\end{aligned}$$

2. **Logistische Regression:** hier gilt $\tau^2 = 1$, $\mu_i = p_i$, $\sigma_i^2 = p_i(1 - p_i)$,

$i = 1, \dots, n$, $p_i \in (0, 1)$ und somit

$$\begin{aligned} U(\beta) &= X^\top(Y - p) , \\ I(\beta) &= X^\top \text{diag}(p_i(1 - p_i))X , \\ W(\beta) &= -I(\beta) , \end{aligned}$$

wobei $p = (p_1, \dots, p_n)^\top$.

3. **Poisson-Regression:** es gilt $\tau^2 = 1$, $\mu_i = \lambda_i = \sigma_i^2$, $i = 1, \dots, n$ und somit

$$\begin{aligned} U(\beta) &= X^\top(Y - \lambda) , \\ I(\beta) &= X^\top \text{diag}(\lambda_i)X , \\ W(\beta) &= -I(\beta) , \end{aligned}$$

wobei $\lambda = (\lambda_1, \dots, \lambda_n)^\top$.

Wann gibt die Lösung des Gleichungssystems $U(\beta) = 0$ einen Maximum-Punkt der Funktion $\log L(Y, \beta)$ an?

Mit anderen Worten: Wann existiert der ML-Schätzer $\hat{\beta}$ von β , der eindeutig bestimmt ist?

$$\hat{\beta} = \underset{\beta}{\operatorname{argmax}} \log L(Y, \beta) \Rightarrow U(\hat{\beta}) = 0$$

Die hinreichenden Bedingungen eines Maximums fordern, dass die Hesse-Matrix $W(\beta)$ dafür negativ definit sein soll.

Betrachten wir den Spezialfall der natürlichen Linkfunktion.

Dann gilt nach Bemerkung 2.6.1:

- Das Gleichungssystem $U(\beta) = 0$ schreibt sich $U(\beta) = \frac{1}{\tau^2} X^\top(Y - \mu(\beta)) = 0$
- Die Matrix $W(\beta) = -\frac{1}{\tau^4} X^\top \text{diag}(\sigma_i^2(\beta))X$ ist negativ definit, falls zusätzlich $\operatorname{rg}(X) = m$ und $0 < \sigma_i^2(\beta) < \infty$ für alle $i = 1, \dots, n$. Unter diesen Bedingungen existiert also eindeutig ein ML-Schätzer $\hat{\beta}$ für β .

Geben wir jetzt Verfahren an, die das (im Allgemeinen nicht lineare) Gleichungssystem $U(\beta) = 0$ numerisch lösen. Diese Ansätze sind (genauso wie in Abschnitt 2.3) iterativ, d.h. sie nähern sich Schrittweise dem ML-Schätzer $\hat{\beta}$ an.

1. Newton-Verfahren

Dieser Ansatz ist dem Gauß-Newton-Verfahren von Abschnitt 2.3 sehr ähnlich:

1. Wähle einen geeigneten Startwert $\hat{\beta}_0 \in \mathbb{R}^m$.

2. Im Schritt $k + 1$, berechne $\hat{\beta}_{k+1}$ aus $\hat{\beta}_k$, $k = 0, 1, \dots$ auf folgende Art und Weise:

- Nehme die Taylor-Entwicklung von $U(\beta)$ bis zur ersten Ordnung an der Stelle

$$\hat{\beta}_k : U(\beta) \approx U(\hat{\beta}_k) + W(\hat{\beta}_k)(\beta - \hat{\beta}_k) \quad (2.6.6)$$

- Setze sie gleich Null: $U(\hat{\beta}_k) + W(\hat{\beta}_k)(\beta - \hat{\beta}_k) = 0$
- Die Lösung dieses Gleichungssystems ist $\hat{\beta}_{k+1}$:

$$\hat{\beta}_{k+1} = \hat{\beta}_k - W^{-1}(\hat{\beta}_k) \cdot U(\hat{\beta}_k) , \quad k = 0, 1, 2, \dots ,$$

vorausgesetzt, dass $W(\hat{\beta}_k)$ invertierbar ist.

3. Breche den Iterationsprozess ab, sobald $|\hat{\beta}_{k+1} - \hat{\beta}_k| < \delta$ für eine vorgegebene Genauigkeit $\delta > 0$ ist.

Das Konvergenzverhalten dieses Verfahrens hängt entscheidend von der Wahl von $\hat{\beta}_0$ ab, für dessen Konvergenz $\hat{\beta}_0$ nah genug bei $\hat{\beta}$ liegen muss. Ein weiterer Nachteil dieses Verfahrens ist, dass die zufällige Matrix $W(\beta)$ unter Umständen nicht invertierbar sein kann. Deswegen schlagen wir jetzt eine Modifikation des Newton-Verfahrens vor, bei der $W(\beta)$ durch den Erwartungswert

$$\mathbb{E} W(\beta) = -I(\beta) \quad (2.6.7)$$

ersetzt wird. Dass die Identität (2.6.6) stimmt, folgt aus dem Satz 2.6.2, Formel (2.6.5) und der Tatsache, dass $\mathbb{E} Y_i = \mu_i$, $i = 1, \dots, n$. Wenn man voraussetzt, dass $rg(X) = m$ und $u_i \neq 0$, $i = 1, \dots, n$, so ist nach Satz 2.6.1 $I(\beta)$ invertierbar. Dieses Verfahren wird *Fisher Scoring* genannt.

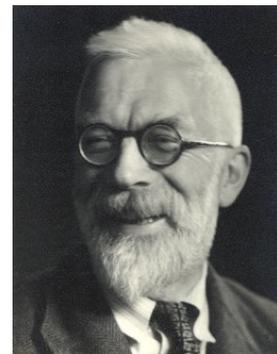
2. Fisher Scoring

Der einzige Unterschied zu den Schritten des Newton-Verfahrens besteht beim Fisher Scoring darin, dass man in Schritt 2 die iterative Gleichung

$$\hat{\beta}_{k+1} = \hat{\beta}_k + I^{-1}(\hat{\beta}_k)U(\hat{\beta}_k) , \quad k = 0, 1, \dots$$

einsetzt.

Im Falle einer natürlichen Linkfunktion gilt nach Bemerkung 2.6.1



R.A. Fisher

$$\begin{aligned} \hat{\beta}_{k+1} &= \hat{\beta}_k + \tau^4 (X^\top \text{diag}(\sigma_i^2(\hat{\beta}_k)) X)^{-1} \frac{1}{\tau^2} (X^\top (Y - \mu(\hat{\beta}_k))) \quad (1890 - 1962) \\ &= \hat{\beta}_k + \tau^2 (X^\top \text{diag}(\sigma_i^2(\hat{\beta}_k)) X)^{-1} (X^\top (Y - \mu(\hat{\beta}_k))) . \quad 1 \end{aligned}$$

¹Die Inverse einer Matrix A multipliziert mit einem Skalar $k \neq 0$ ist $(kA)^{-1} = k^{-1}A^{-1}$

2.6.2 Asymptotische Tests für β

Das Ziel dieses Abschnittes ist es, eine Testregel für die Hypothese

$$H_0 : \beta = \beta_0 \text{ vs. } H_1 : \beta \neq \beta_0 \quad \text{mit} \quad \beta = (\beta_1, \dots, \beta_m)^\top, \quad \beta_0 = (\beta_{01}, \dots, \beta_{0m})^\top$$

zu konstruieren. Insbesondere sind die Haupthypothesen $H_0 : \beta = 0$ bzw. $H_0 : \beta_j = 0$ von Interesse, weil sie die Tatsache reflektieren, dass die Zielvariablen $Y = (Y_1, \dots, Y_n)^\top$ von einigen Ausgangsvariablen (z.B. $(x_{1j}, \dots, x_{nj})^\top$ im Falle der Hypothese $\beta_j = 0$) unabhängig sind.

Um solche Hypothesen testen zu können, werden Teststatistiken T_n vorgeschlagen, die asymptotisch (für $n \rightarrow \infty$) eine bekannte Prüfverteilung (z.B. multivariate Normalverteilung oder χ^2 -Verteilung) besitzen. Dafür sind gewisse Vorarbeiten notwendig.

Definition 2.6.5

Sei $\{Z_n\}_{n \in \mathbb{N}}$ eine Folge von k -dimensionalen Zufallsvektoren. Man sagt, dass

1. $Z_n \xrightarrow[n \rightarrow \infty]{d} Z$ (Konvergenz in Verteilung gegen einen Zufallsvektor Z), falls $P(Z_n \leq x) \xrightarrow[n \rightarrow \infty]{d} P(Z \leq x) \quad \forall x = (x_1, \dots, x_n)^\top \in \mathbb{R}^n$, wobei $\{Z_n \leq x\}$ bedeutet, dass $\{Z_{nj} \leq x_j, \quad j = 1, \dots, k\}$ für $Z_n = (Z_{n1}, \dots, Z_{nk})^\top$.
2. $Z_n \xrightarrow[n \rightarrow \infty]{P} Z$ (Konvergenz in Wahrscheinlichkeit gegen einen Zufallsvektor Z), falls $P(|Z_n - Z| > \varepsilon) \xrightarrow[n \rightarrow \infty]{} 0 \quad \forall \varepsilon > 0$, wobei $|Z_n - Z|$ die Euklidische Norm des Vektors $Z_n - Z$ bezeichnet.

Sei

$$g(\mathbb{E} Y_i) = X_i \beta, \quad i = 1, \dots, n,$$

ein verallgemeinertes lineares Modell mit natürlicher Linkfunktion g . Seien $L(Y, \beta)$, $U(\beta)$ und $I(\beta)$ die Likelihood-Funktion bzw. der Vektor der partiellen Ableitungen von $\log L(Y, \beta)$ bzw. die Fisher-Informationsmatrix in diesem Modell.

Es gelten folgende Voraussetzungen:

1. $\hat{\beta}_n = \hat{\beta}(Y_1, \dots, Y_n, X)$ sei eine schwachkonsistente Folge von ML-Schätzern für β : $U(\hat{\beta}_n) = 0$, $n \in \mathbb{N}$ und $\hat{\beta}_n \xrightarrow[n \rightarrow \infty]{P} \beta$.
2. Es existiert eine Folge $\{\Gamma_n\}_{n \in \mathbb{N}}$ von invertierbaren $(m \times m)$ -Matrizen $\Gamma_n = \Gamma_n(\beta)$ mit den Eigenschaften $\lim_{n \rightarrow \infty} \Gamma_n = 0$ und $\lim_{n \rightarrow \infty} \Gamma_n^\top I_n(\beta) \Gamma_n = K^{-1}(\beta)$, wobei $K(\beta)$ eine symmetrische positiv definite $(m \times m)$ -Matrix ist.

Satz 2.6.3

Unter obigen Voraussetzungen gilt:

1. $T_n^* = \Gamma_n^{-1}(\hat{\beta}_n - \beta) \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}(0, K(\beta))$, wobei $K^{-1}(\beta) = \Gamma_n^\top I_n(\beta) \Gamma_n$ und
2. $T_n = 2(\log L(Y, \hat{\beta}_n) - \log L(Y, \beta)) \xrightarrow[n \rightarrow \infty]{d} \chi_m^2$, $m = \dim \beta$

Bemerkung 2.6.2

1. Falls $I_n(\beta)$ positiv definit ist und $\lim_{n \rightarrow \infty} I_n^{-1}(\beta) = 0$ (die Konvergenz versteht sich elementweise), dann kann $\Gamma_n = (I_n^{-\frac{1}{2}})^\top(\beta)$ gewählt werden. ²

Bei dieser Auswahl von Γ_n ist die asymptotische Kovarianzmatrix $K(\beta)$ gleich $\text{Id} = \text{diag}(1, \dots, 1)$:

$$K^{-1}(\beta) = \Gamma_n^\top I_n(\beta) \Gamma_n = I_n^{-\frac{1}{2}}(\beta) I_n^{\frac{1}{2}}(\beta) (I_n^{\frac{1}{2}})^\top(\beta) (I_n^{-\frac{1}{2}})^\top(\beta) = \text{Id} .$$

In diesem Fall lautet die Aussage 1 des Satzes 2.6.3

$$(I_n^{\frac{1}{2}})^\top(\beta) (\hat{\beta}_n - \beta) \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}(0, \text{Id}) . \quad (2.6.8)$$

2. Betrachten wir den Spezialfall des logistischen Regressionsmodells:

$Y_i \sim \text{Bernoulli}(p_i)$, $i = 1, \dots, n$.

$$g(\mathbb{E} Y_i) = \log \frac{P(Y_i = 1)}{P(Y_i = 0)} = \sum_{j=1}^m \beta_j x_{ij} \Leftrightarrow \log \frac{p_i}{1 - p_i} = \sum_{j=1}^m \beta_j x_{ij}, \quad i = 1, \dots, n .$$

Aus Beispiel 2.6.3(2) ist

$$I_n(\beta) = X^\top \text{diag}(p_i(1 - p_i)) X, \quad \text{wobei } X = (x_{ij})_{\substack{i=1, \dots, n \\ j=1, \dots, m}} \text{ unsere Designmatrix ist.}$$

Falls $p_i \in (0, 1)$ und $\inf_i p_i(1 - p_i) > 0$, $\text{rg}(X) = m$, dann ist $I_n(\beta)$ positiv definit. Falls zusätzlich $\lim_{n \rightarrow \infty} (X^\top X)^{-1} = 0$, dann gilt $\lim_{n \rightarrow \infty} I_n^{-1}(\beta) = 0$ und somit (2.6.7).

Wie verwendet man nun den Satz 2.6.3 zum Testen der Hypothesen

$$H_0 : \beta = \beta_0 \quad \text{vs.} \quad H_1 : \beta \neq \beta_0 ,$$

oder komponentenweise

$$H_0 : \beta_j = \beta_{j0}, \quad j = 1, \dots, m \quad \text{vs.} \quad H_1 : \exists j_1 : \beta_{j_1} \neq \beta_{j_10} \quad ?$$

Sei

$$g(\mathbb{E} Y_i) = \sum_{j=1}^m x_{ij} \beta_j, \quad i = 1, \dots, n ,$$

ein verallgemeinertes lineares Modell mit natürlicher Linkfunktion g .

²Die quadratische Wurzel aus einer Matrix A : falls A eine positiv definite Matrix ist, dann gibt es eine Matrix $A^{\frac{1}{2}} : A = A^{\frac{1}{2}} (A^{\frac{1}{2}})^\top$

Nach Bemerkung 2.6.1 gilt

$$\log L(Y, \beta) = \frac{1}{\tau^2} \sum_{i=1}^n \left(Y_i x_i^\top \beta + a(Y_i, \tau) - b(x_i^\top \beta) \right), \quad Y = (Y_1, \dots, Y_n), \quad x_i = (x_{i1}, \dots, x_{im}).$$

Deshalb gilt

$$T_n = \frac{2}{\tau^2} \sum_{i=1}^n \left(Y_i x_i^\top (\hat{\beta}_n - \beta_0) - b(x_i^\top \hat{\beta}_n) + b(x_i^\top \beta_0) \right)$$

Bei Vorgabe eines Exponential-Modells (τ, b - bekannt), der Stichprobe der Zielvariablen Y und der Designmatrix X wird H_0 verworfen, falls $T_n \geq \chi_{m,1-\alpha}^2$, wobei m die Anzahl der Parameter im Modell, $\chi_{m,1-\alpha}^2$ das $(1 - \alpha)$ -Quantil der χ_m^2 - Verteilung und $\alpha \in (0, 1)$ das Signifikanzniveau des asymptotischen Tests ist. Dieser Test ist nur für relativ große n anwendbar. Der Fehler 1. Art hat dabei (für $n \rightarrow \infty$) die asymptotische Wahrscheinlichkeit α . Falls eine einfache Hypothese

$$H_0 : \beta_j = 0 \text{ vs. } H_1 : \beta_j \neq 0$$

getestet werden soll, benutzt man die Teststatistik T_n^* . H_0 wird verworfen, falls

$$|T_n^*| = \frac{|\hat{\beta}_{nj}|}{\sqrt{(I_n^{-1}(\hat{\beta}_n))_{jj}}} > z_{1-\frac{\alpha}{2}},$$

wobei $z_{1-\frac{\alpha}{2}}$ das $(1 - \alpha)$ -Quantil der $\mathcal{N}(0, 1)$ - Verteilung ist. Im Nenner steht eine Schätzung der asymptotischen Standardabweichung von $\hat{\beta}_{kj}$. Dies ist ein asymptotischer Test zum Niveau α , weil

$$\begin{aligned} P_{H_0}(|T_n^*| > z_{1-\frac{\alpha}{2}}) &= 1 - P_{H_0}(|T_n^*| \leq z_{1-\frac{\alpha}{2}}) \xrightarrow{n \rightarrow \infty} 1 - \Phi(z_{1-\frac{\alpha}{2}}) + \underbrace{\Phi(-z_{1-\frac{\alpha}{2}})}_{1-\Phi(z_{1-\frac{\alpha}{2}})} \\ &= 1 - \left(1 - \frac{\alpha}{2}\right) + 1 - \left(1 - \frac{\alpha}{2}\right) = \alpha, \end{aligned}$$

wobei

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt$$

die Verteilungsfunktion der $\mathcal{N}(0, 1)$ - Verteilung ist.

Beispiel 2.6.4 (Kreditrisikoprüfung)

vgl. Fahrmeir, L., Kneib, T., Lang, S. - Regression, S.208ff

Es liegt folgender Datensatz einer süddeutschen Bank aus den 1990er Jahren vor:

Es werden Ergebnisse der Kreditrisikoprüfung von $n = 1000$ Kreditanträgen (ca. 700 gute und 300 schlechte Kredite) analysiert:

$$\text{Zielvariable } Y_i = \begin{cases} 0, & \text{falls das Darlehen vom Kunden } i \text{ zurückgezahlt wurde} \\ 1, & \text{falls das Darlehen vom Kunden } i \text{ nicht zurückgezahlt wurde} \end{cases}$$

Die Designmatrix X enthält folgende Zusatzinformationen über den Kunden:

$$x_{i1} - \text{Kontoführung des Kontos bei der Bank:} = \begin{cases} 1, & \text{kein Konto} \\ 0, & \text{sonst} \end{cases}$$

$$x_{i2} - \text{Bewertung der Kontoführung:} = \begin{cases} 1, & \text{gutes Konto} \\ 0, & \text{kein oder schwaches Konto} \end{cases}$$

x_{i3} - Laufzeit des Kredits in Monaten

x_{i4} - Höhe des Kredits in DM

$$x_{i5} - \text{Zahlungsverhalten beim Kunden :} = \begin{cases} 1, & \text{gut} \\ 0, & \text{sonst} \end{cases}$$

$$x_{i6} - \text{Verwendungszweck:} = \begin{cases} 1, & \text{privat} \\ 0, & \text{geschäftlich} \end{cases}$$

		$Y = 1$	$Y = 0$
x_1	kein Konto	45.0	20.0
x_2	gut	15.3	49.8
	schlecht	39.7	30.2
x_4	Kredithöhe	$Y = 1$	$Y = 0$
	$0 < \dots \leq 500$	1.00	2.14
	$500 < \dots \leq 1000$	11.33	9.14
	$1000 < \dots \leq 1500$	17.00	19.86
	$1500 < \dots \leq 2500$	19.67	24.57
	$2500 < \dots \leq 5000$	25.00	28.57
	$5000 < \dots \leq 7500$	11.33	9.71
	$7500 < \dots \leq 10000$	6.67	3.71
	$10000 < \dots \leq 15000$	7.00	2.00
	$15000 < \dots \leq 20000$	1.00	0.29
x_5	Frühere Kredite	$Y = 1$	$Y = 0$
	gut	82.33	94.95
	schlecht	17.66	5.15
x_6	Verwendungszweck	$Y = 1$	$Y = 0$
	privat	57.53	69.29
	beruflich	42.47	30.71

Tabelle 2.6.4a - Auszug aus dem Originaldatensatz

\bar{x}_1	\bar{x}_2	\bar{x}_3	\bar{x}_4	\bar{x}_5	\bar{x}_6
0.274	0.393	20.903	3271	0.911	0.657

Tabelle 2.6.4b - Mittelwerte \bar{x}_j von x_{ij} im Datensatz

Frage: Wie soll $\hat{\beta}$ geschätzt werden?

Als Modell wird das Logit-Modell gewählt mit $p_i = P(Y_i = 1)$, $i = 1, \dots, n$:

$$\log \frac{p_i}{1 - p_i} = \beta_0 + x_{i1}\beta_1 + x_{i2}\beta_2 + x_{i3}\beta_3 + x_{i4}\beta_4 + x_{i5}\beta_5 + x_{i6}\beta_6 \quad \text{für } i = 1, \dots, n,$$

wobei $\beta = (\beta_0, \dots, \beta_6)^\top$, $m = 7$.

Ziel: Schätze β_0, \dots, β_6 und prüfe, welche Faktoren für die künftige Kreditvergabe relevant sind.

	Wert	$\sqrt{(I_n^{-1}(\hat{\beta}))_{ii}}$	T_n^*	p-Wert
β_0	0.281	0.303	-0.94	0.347
β_1	0.618	0.175	3.53	< 0.001
β_2	-1.338	0.201	-6.65	< 0.001
β_3	0.033	0.008	4.29	< 0.001
β_4	0.023	0.033	0.72	0.474
β_5	-0.986	0.251	-3.93	< 0.001
β_6	-0.426	0.266	-2.69	0.007

Tabelle 2.6.4c - Ergebnis zur ML-Schätzung durch das Fisher Scoring Verfahren

Wobei $\sqrt{(I_n^{-1}(\hat{\beta}))_{ii}}$ die asymptotische Standardabweichung von $\hat{\beta}_i$ ist (vgl. Satz 2.6.3).
Signifikanzniveau: $\alpha = 0.001$

$H_0 : \beta_i = 0$ (Merkmal x_i beeinflusst die Kreditvergabe nicht) wird abgelehnt, falls p-Wert $\leq \alpha$. Man sieht, dass u.a. auch β_4 für die Kreditvergabe nicht relevant ist, was der Intuition widerspricht. \Rightarrow Eine Verfeinerung des Modells ist notwendig:

Neues Modell:

$$g(\mathbb{E} Y_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3^1 x_{i3} + \beta_3^2 x_{i3}^2 + \beta_4^1 x_{i4} + \beta_4^2 x_{i4}^2 + \beta_5 x_{i5} + \beta_6 x_{i6}$$

	Wert	$\sqrt{(I_n^{-1}(\hat{\beta}))_{ii}}$	T_n^*	p-Wert
β_0	-0.488	0.390	-1.25	0.211
β_1	0.618	0.176	3.51	< 0.001
β_2	-1.337	0.202	-6.61	< 0.001
β_3^1	0.092	0.025	3.64	< 0.001
β_3^2	-0.001	< 0.001	-2.20	0.028
β_4^1	-0.264	0.099	-2.68	0.007
β_4^2	0.023	0.007	3.07	0.002
β_5	-0.995	0.255	-3.90	< 0.001
β_6	-0.404	0.160	-2.52	0.012

Tabelle 2.6.4d

Frage: Welches Modell ist besser?

Mit anderen Worten, wir testen

$$H_0 : \beta_3^2 = 0 \text{ (lineares Modell) vs. } H_1 : \beta_3^2 \neq 0 \text{ (quadratisches Modell) bzw.}$$

$$H_0 : \beta_4^2 = 0 \text{ (lineares Modell) vs. } H_1 : \beta_4^2 \neq 0 \text{ (quadratisches Modell) .}$$

Dabei verallgemeinern wir die Art der statistischen Hypothesen wie folgt: es wird

$$H_0 : C\beta = d \text{ vs. } H_1 : C\beta \neq d$$

getestet, wobei C eine $(r \times m)$ - Matrix mit $rg C = r \leq m$ ist und $d \in \mathbb{R}^r$.

Zum Vergleich: früher haben wir

$$H_0 : \beta = \beta_0 \text{ vs. } H_1 : \beta \neq \beta_0, \quad \beta, \beta_0 \in \mathbb{R}^m$$

getestet. Natürlich ist $\beta = \beta_0$ ein Spezialfall von $C\beta = d$ mit $C = \text{Id}$, $d = \beta_0$. Die neuen Hypothesen beinhalten Aussagen über die Linearkombinationen der Parameterwerte. Wie soll H_0 vs. H_1 getestet werden?

Sei $\tilde{\beta}$ der ML-Schätzer von β unter H_0 , d.h. $\tilde{\beta} = \underset{\beta \in \mathbb{R}^m: C\beta=d}{\operatorname{argmax}} \log L(Y, \beta)$

Sei $\hat{\beta}$ der ML-Schätzer von β unrestringiert, d.h. $\hat{\beta} = \underset{\beta \in \mathbb{R}^m}{\operatorname{argmax}} \log L(Y, \beta)$.

Die Idee der folgenden Tests ist es, $\tilde{\beta}$ mit $\hat{\beta}$ zu vergleichen. Falls die Abweichung $\hat{\beta} - \tilde{\beta}$ groß ist, soll H_0 abgelehnt werden.

Satz 2.6.4

Sei $\log L(Y, \beta)$ die Log-Likelihood-Funktion der Stichprobe der Zielvariablen $Y = (Y_1, \dots, Y_n)^\top$, $I_n(\beta)$ die Fisher-Informationsmatrix, $U(\beta)$ die Score-Funktion des verallgemeinerten linearen Modells mit natürlicher Linkfunktion

$$g : g(\mathbb{E} Y_i) = X_i \beta, \quad i = 1, \dots, n.$$

Wir führen folgende Teststatistiken ein:

1. Likelihood-Ratio-Teststatistik:

$$\tilde{T}_n = 2(\log L(Y, \hat{\beta}) - \log L(Y, \tilde{\beta}))$$

2. Wald-Statistik:

$$\tilde{T}_n^* = (C\hat{\beta} - d)^\top (CI_n^{-1}C^\top)^{-1} (C\hat{\beta} - d)$$

3. Score-Statistik:

$$\bar{T}_n^* = U(\tilde{\beta})^\top I_n^{-1}(\tilde{\beta}) U(\tilde{\beta})$$

Unter gewissen Bedingungen an die Schätzer $\hat{\beta}$ und $\tilde{\beta}$ (vgl. Satz 2.6.3) sind die Teststatistiken 1 - 3 asymptotisch χ_m^2 -verteilt: z.B. gilt für die Likelihood-Quotienten-Teststatistik

$$\tilde{T}_n \xrightarrow[n \rightarrow \infty]{d} \chi_m^2.$$

Folgerung 2.6.1

Der Satz 2.6.4 liefert uns folgende Entscheidungsregel: H_0 wird abgelehnt, falls

$$\tilde{T}_n(\tilde{T}_n^*, \bar{T}_n) > \chi_{m, 1-\alpha}^2.$$

Dies ist ein asymptotischer Test zum Signifikanzniveau α .

Beispiel 2.6.4 (Fortsetzung)

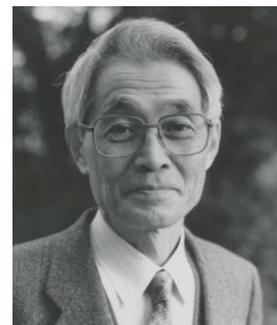
Es ergeben sich folgende Werte für die Teststatistiken:

$$\begin{aligned} \tilde{T}_n &= 12.44, & \text{p-Wert: } 0.0020 \\ \tilde{T}_n^* &= 11.47, & \text{p-Wert: } 0.0032. \end{aligned}$$

Für $\alpha = 0.005$ gilt p-Wert $\leq \alpha$, somit wird $H_0 : \beta_4^2 = 0$ abgelehnt \Rightarrow das quadratische verallgemeinerte lineare Modell ist besser.

2.6.3 Kriterien zur Modellwahl bzw. Modellanpassung

Es ist bekannt, dass die Güte der Anpassung eines parametrischen Modells an die Daten im Allgemeinen steigt, wenn die Anzahl der Parameter erhöht wird. Die Aufgabe eines Ökonometrikers ist es aber ein gut passendes Modell mit einer möglichst kleinen Anzahl an Parametern zu finden. Deshalb verwendet man folgendes Informationskriterium von Akaike, um Modelle mit (möglicherweise) unterschiedlichen Parametersätzen zu vergleichen.



H. Akaike
(1927 -)

Informationskoeffizient von Akaike:

$$\text{AIC} = -2 \log L(Y, \hat{\beta}) + 2m ,$$

wobei $Y = (Y_1, \dots, Y_n)$ die Stichprobe der Zielvariablen im verallgemeinerten linearen Modell und $\hat{\beta}$ der dazugehörige ML-Schätzer sei. Der Wert von AIC berücksichtigt einerseits die Forderung der Maximalität der Log-Likelihood-Funktion $\log L(Y, \hat{\beta})$, andererseits bestraft er Modelle mit einer großen Anzahl von Parametern m . Das Modell mit dem kleineren AIC ist als besseres Modell einzustufen. Manchmal verwendet man statt AIC den normierten Koeffizienten AIC/n .

Beispiel 2.6.4 (Fortsetzung)

Berechnen wir den Informationskoeffizienten von Akaike für das lineare und quadratische Logit-Modell im Beispiel der Kreditrisikoprüfung:

$$\text{Lineares Modell : AIC} = 1043.815$$

$$\text{Quadratisches Modell : AIC} = 1035.371$$

Man sieht anhand des AIC, dass die Wahl zu Gunsten des quadratischen Modells ausfällt.

Der Nachteil der oben beschriebenen AIC-Regel liegt darin, dass die endgültige Entscheidung dem Ökonometriker überlassen bleibt. Deshalb ist es wünschenswert, einen statistischen Test zu konstruieren, der die Güte der Modellanpassung beurteilen kann.

Wir werden jetzt den χ^2 -Test beschreiben.

Sei

$$g(\mathbb{E} Y_i) = X_i \beta , \quad i = 1, \dots, n ,$$

ein verallgemeinertes lineares Modell mit Linkfunktion g und Parametervektor $\beta = (\beta_1, \dots, \beta_m)^\top$.

Teilen wir die Zielvariablen Y_1, \dots, Y_n in k Gruppen auf, so dass sie möglichst homogen in Bezug auf die zu schätzenden Parametern sind. So liegt z.B. eine solche Aufteilung vor, wenn der Wertebereich der Zielvariablen Y_i „geschickt“ in k Intervalle $(a_l, b_l]$ unterteilt wird:

$$-\infty \leq a_1 < b_1 = a_2 < b_2 = a_3 < \dots < b_{k-1} = a_k < b_k \leq +\infty$$

In die Gruppe l fallen alle Beobachtungen Y_i , die zu $(a_l, b_l]$ gehören. Dabei müssen $(a_l, b_l]$ so gewählt werden, dass $\hat{\mu}_j = g^{-1}(X_j \hat{\beta})$ innerhalb einer Gruppe konstant wird: $\hat{\mu}_j \equiv \hat{\mu}_l \forall j$ aus Gruppe l .³ Sei

- $n_l = \# \{Y_j : Y_j \in (a_l, b_l]\}$ die Klassenstärke der Klasse l

³Dies ist eine informelle Beschreibung des Vorgangs, bei dem für jedes Y_i n_i unabhängige Kopien von Y_i erzeugt werden, die die i -te Klasse bilden.

- $\bar{Y}_l = \frac{1}{n_l} \sum Y_j$ das arithmetische Mittel innerhalb der Klasse l
- $\hat{\beta}$ der ML-Schätzer von β , der aus Y gewonnen wurde
- $l_l(\beta) = \sum \log f_\theta(Y_j)$ die Log-Likelihood-Funktion der Zielvariablen Y_i innerhalb der Gruppe l
- $\hat{\mu}_l = g^{-1}(X_l \hat{\beta})$ und $v(\hat{\mu}_l)$ der Erwartungswert- bzw. der Varianzschätzer von $\mu_l = \mathbb{E} Y_l$, die aus dem ML-Schätzer $\hat{\beta}$ gewonnen wurden

Dabei ist $v(\hat{\mu}_l) = \tau^2 b''(b'^{-1}(\hat{\mu}_l))$, wobei $b(\cdot)$ der entsprechende Koeffizient in der Dichte f_θ aus der Exponentialfamilie ist. Man bildet folgende Teststatistiken:

$$\chi^2 = \sum_{l=1}^k \frac{(\bar{Y}_l - \hat{\mu}_l)^2}{v(\hat{\mu}_l)/n_l}$$

$$D = -2\tau^2 \sum_{l=1}^k (l_l(\hat{\mu}_l) - l_l(\bar{Y}_l))$$

Satz 2.6.5

Falls $n \rightarrow \infty$ und die Anzahl $n_l \rightarrow \infty \forall l$, dann gilt unter gewissen Voraussetzungen Folgendes:

$$\chi^2 \xrightarrow[n \rightarrow \infty]{d} \chi_{k-m-1}^2$$

$$D \xrightarrow[n \rightarrow \infty]{d} \chi_{k-m-1}^2$$

Folgerung 2.6.2

Mit Hilfe der Behauptungen des Satzes 2.6.5 können die Hypothesen

$$H_0 : Y = (Y_1, \dots, Y_n) \text{ stammt aus dem Modell } g(\mathbb{E} Y_i) = X_i \beta, \quad i = 1, \dots, n$$

vs.

$$H_1 : Y = (Y_1, \dots, Y_n) \text{ stammt nicht aus dem Modell } g(\mathbb{E} Y_i) = X_i \beta, \quad i = 1, \dots, n$$

folgendermaßen getestet werden:

H_0 wird (für große n) zum asymptotischen Signifikanzniveau α verworfen, falls

$$\chi^2 > \chi_{k-m-1, 1-\alpha}^2 \quad \text{bzw.} \quad D > \chi_{k-m-1, 1-\alpha}^2 .$$

Diese Tests sollten aber nicht verwendet werden, falls die Klassenstärken n_l klein sind.

Beispiel 2.6.5

Wie sehen die oben beschriebenen Tests im Falle der Logit- bzw. Poisson-Regression aus?

1. **Logit-Modell:** $Y_i \sim \text{Bernoulli}(p_i)$, $i = 1, \dots, n$

$$\Rightarrow \text{verallgemeinertes lineares Modell} \quad \log \frac{p_i}{1-p_i} = X_i \beta, \quad i = 1, \dots, n$$

Wir teilen Y_1, \dots, Y_n in k Klassen auf, so dass die Wahrscheinlichkeit des Auftretens von 1 in jeder Klasse möglichst gut durch $\bar{Y}_l = \frac{1}{n_l} \sum Y_i$ geschätzt wird. Somit gilt mit $\hat{\mu}_l = \hat{p}_l = g^{-1}(X_l \hat{\beta}) = \frac{e^{X_l \hat{\beta}}}{1+e^{X_l \hat{\beta}}}$, $v(\hat{p}_l) = \hat{p}_l(1 - \hat{p}_l)$

$$\Rightarrow \chi^2 = \sum_{l=1}^k \frac{(\bar{Y}_l - \hat{p}_l)^2}{\hat{p}_l(1 - \hat{p}_l)/n_l}$$

2. **Poisson-Modell:** $Y_i \sim \text{Poisson}(\lambda)$,

$$\Rightarrow \text{verallgemeinertes lineares Modell} \quad \log \lambda_i = X_i \beta, \quad i = 1, \dots, n$$

Somit gilt mit $\hat{\mu}_l = \hat{\lambda}_l = e^{X_l \hat{\beta}}$, $v(\hat{\lambda}_l) = \hat{\lambda}_l$

$$\Rightarrow \chi^2 = \sum_{l=1}^k \frac{(\bar{Y}_l - \hat{\lambda}_l)^2}{\hat{\lambda}_l/n_l}$$

3 Zeitreihenanalyse

3.1 Korrelierte Beobachtungen im Regressionsmodell

Sehr oft (insbesondere im Falle der Zeitreihen, also zeitlich aufgelösten Beobachtungen) ist die Unabhängigkeit der beobachtbaren Größen nicht mehr gegeben (z.B. zeitlicher Ablauf von Aktienkursen o.Ä.).

Beispiel 3.1.1 (Absatzzahlen von Wasserfiltern)

vgl. v. Auer - Ökonometrie, S. 389ff

t	x_t	Y_t	t	x_t	Y_t
1	24.2	1990	13	32.2	1700
2	25.5	1630	14	32.4	1450
3	26.8	1570	15	33.2	1480
4	26.4	1960	16	34.0	1450
5	25.2	2150	17	33.7	1000
6	24.4	2450	18	32.8	1080
7	26.2	2210	19	31.3	1270
8	26.1	2400	20	30.9	1520
9	27.4	2200	21	30.0	1820
10	28.4	1270	22	28.3	1660
11	29.8	1250	23	27.5	1500
12	31.3	1500	24	26.8	1810

Tabelle 3.1.1

Es sind Absatzmengen Y_t von Wasserfiltern in 1000 Stück eines Marktführers in Abhängigkeit vom Verkaufspreis x_t (in Euro) für den Zeitraum Januar 2005 bis Dezember 2006 aufgelöst nach Monaten ($t = 1, \dots, 24$) zu analysieren. Es wird versucht, ein Modell der einfachen linearen Regression

$$Y_t = \alpha + \beta x_t + \varepsilon_t, \quad t = 1, \dots, 24$$

anzupassen. Die einfache lineare Regression setzt voraus, dass die Störgrößen ε_t (und somit auch Y_t) unabhängig oder zumindest unkorreliert sind. Obwohl die Berechnung der MKQ-Schätzer $\hat{\alpha}$ und $\hat{\beta}$ für α und β keiner Annahmen bedarf, hängen die Eigenschaften von $\hat{\alpha}$ und $\hat{\beta}$ entscheidend von der Annahme der Unabhängigkeit und weiteren Verteilungsannahmen ab. Dies ermöglicht auch die Beurteilung der Güte der Modellanpassung durch die entsprechenden Tests. Wie wir gleich sehen werden, ist die Annahme der Unkorreliertheit in diesem Beispiel verletzt, was es uns unmöglich macht zu beurteilen, wie

gut unser Modell die Daten darstellt: Obwohl die MKQ-Schätzer

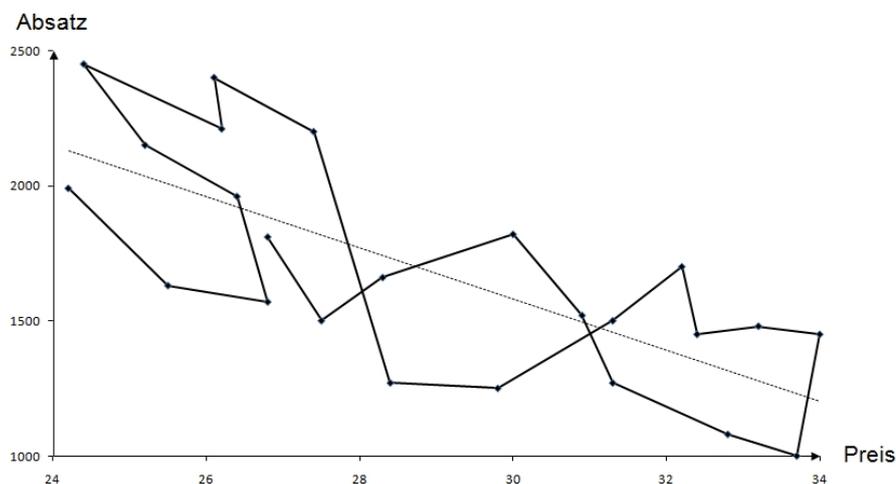
$$\hat{\alpha} = 4413.33, \quad \hat{\beta} = -94.42$$

leicht zu berechnen sind, sind sie nutzlos, falls weitere Güteaussagen fehlen.

Bei der grafischen Darstellung der Punktwolke (x_t, Y_t) , $t = 1, \dots, 24$, fällt Folgendes auf: Ein lineares Modell scheint auf den ersten Blick geeignet zu sein, um die Daten zu beschreiben, weil die Punkte (x_t, Y_t) relativ gleichmäßig um die Regressionsgerade

$$y = \hat{\alpha} + \hat{\beta}x$$

streuen. Dabei fällt auf, dass $\hat{\beta} < 0$, was der natürlichen Tatsache entspricht, dass mit steigendem Preis die Nachfrage sinkt.



Beim Verbinden der aufeinanderfolgenden Punkte in ihrer zeitlichen Reihenfolge wird aber erkennbar, dass Y_t positiv korreliert sind, weil die Verbindungslinie die Regressionsgerade nicht so oft schneidet. So bedeutet eine positive Abweichung von der Regressionsgeraden im Monat t mit hoher Wahrscheinlichkeit weitere positive Abweichungen in den Folgemonaten $t + 1$, $t + 2, \dots$.

Deshalb stellt man den Bedarf fest, die Modellvorstellungen der Regression auf korrelierte Zielvariablen Y_t zu erweitern. Dies gelingt bei der Betrachtung zeitlicher Abläufe wie im Beispiel 3.1.1 mit Hilfe spezieller stochastischer Prozesse, die *Zeitreihen* genannt werden.

3.2 Zeitreihen

Für die Störgrößen im Beispiel 3.1.1 schlagen wir folgendes Modell vor:

$$\varepsilon_t = \rho\varepsilon_{t-1} + \delta_t, \quad t \in \mathbb{N}, \quad \varepsilon_0 = 0, \quad \{\delta_t\} \text{ u.i.v.}, \quad |\rho| < 1.$$

Somit wird die Störgröße ε_{t-1} in abgeschwächter Form (mit dem Faktor $|\rho| < 1$) im nächsten Schritt weitergeben + unabhängiger zufälliger Effekt δ_t .

$\{\varepsilon_t\}_{t \in \mathbb{N}}$ als stochastischer Prozess besitzt einen Namen: der *autoregressive Prozess 1. Ordnung*.

3.2.1 Autoregressive Prozesse: eine Einführung

Definition 3.2.1

Sei $\{\delta_t\}_{t \in \mathbb{Z}}$ eine Folge von unabhängigen identisch verteilten Zufallsvariablen,

$$\delta_t \sim \mathcal{N}(0, \sigma_\delta^2).$$

Der stochastische Prozess in diskreter Zeit $Z = \{Z_t\}_{t \in \mathbb{Z}}$ heißt *stationärer autoregressiver Prozess 1. Ordnung* (AR(1)), falls $\exists \rho \in (-1; 1)$:

$$Z_t = \rho Z_{t-1} + \delta_t, \quad t \in \mathbb{Z}.$$

Bemerkung 3.2.1

Analog lässt sich ein AR(p)-Prozess Z (autoregressiver Prozess der Ordnung $p \geq 1$) definieren:

$$Z_t = \rho_1 Z_{t-1} + \rho_2 Z_{t-2} + \dots + \rho_p Z_{t-p} + \delta_t, \quad t \in \mathbb{Z} \text{ für } \rho_1, \dots, \rho_p \in (-1; 1).$$

Satz 3.2.1 (Eigenschaften des AR(1)-Prozesses)

Sei $Z = \{Z_t\}_{t \in \mathbb{Z}}$ ein AR(1)-Prozess mit dem Abhängigkeitsfaktor $\rho \in (-1; 1)$. Dann gelten folgende Eigenschaften:

1. $Z_t = \sum_{j=0}^{\infty} \rho^j \delta_{t-j}$
2. $\mathbb{E} Z_t = 0$, $\sigma^2 := \text{Var}(Z_t) = \frac{\sigma_\delta^2}{1-\rho^2} \geq \sigma_\delta^2$
3. $\text{Cor}(Z_t, Z_{t-k}) = \rho^k \sigma^2$, $k \in \mathbb{N}$. Insbesondere gilt $\text{Cor}(Z_t, Z_{t-1}) = \rho$.

Somit liegt bei ρ

$\left\{ \begin{array}{l} > 0 \\ = 0 \\ < 0 \end{array} \right.$	$\left. \begin{array}{l} \text{positive} \\ \text{keine} \\ \text{negative} \end{array} \right\}$	Korrelation zwischen Z_t und Z_{t-k} vor.
--	---	---

Beweis

1. Es gilt

$$\begin{aligned} Z_t &= \rho Z_{t-1} + \delta_t = \rho(\rho Z_{t-2} + \delta_{t-1}) + \delta_t = \rho^2 Z_{t-2} + \rho \delta_{t-1} + \delta_t \\ &= \rho^2(\rho Z_{t-3} + \delta_{t-2}) + \rho \delta_{t-1} + \delta_t = \dots = \sum_{j=0}^{\infty} \rho^j \delta_{t-j} . \end{aligned}$$

2. Es gilt

$$\mathbb{E} Z_t = \mathbb{E} \sum_{j=0}^{\infty} \rho^j \delta_{t-j} = \sum_{j=0}^{\infty} \rho^j \underbrace{\mathbb{E} \delta_{t-j}}_{=0} = 0 ,$$

weil man die Reihenfolge der Summe und des Erwartungswertes vertauschen kann. Warum dies möglich ist, wird im Beweis des Satzes 3.2.2 ausführlich erklärt.

$$\begin{aligned} \text{Var}(Z_t) &= \mathbb{E} Z_t^2 - \underbrace{(\mathbb{E} Z_t)^2}_{=0} = \mathbb{E}(Z_t^2) = \mathbb{E} \left(\sum_{i=0}^{\infty} (\rho^i \delta_{t-i})^2 \right) \\ &= \mathbb{E} \left(\sum_{i,j=0}^{\infty} \rho^{i+j} \delta_{t-j} \delta_{t-i} \right) \\ &= \sum_{i,j=0}^{\infty} \rho^{i+j} \mathbb{E}(\delta_{t-j} \delta_{t-i}) \\ &= \sum_{i,j=0; i \neq j}^{\infty} \rho^{i+j} \underbrace{\mathbb{E} \delta_{t-j}}_{=0} \underbrace{\mathbb{E} \delta_{t-i}}_{=0} + \sum_{i=0}^{\infty} \rho^{2i} \mathbb{E} \delta_{t-i}^2 \\ &= \sum_{i=0}^{\infty} \rho^{2i} \text{Var}(\delta_{t-i}) = \sigma_{\delta}^2 \sum_{i=0}^{\infty} \rho^{2i} = \frac{\sigma_{\delta}^2}{1 - \rho^2} \geq \sigma_{\delta}^2 . \end{aligned}$$

3. Beweisen wir den Spezialfall $k = 1$. Der allgemeine Fall $k > 1$ bleibt dem Leser als Übungsaufgabe überlassen.

$$\begin{aligned} \text{Cov}(Z_t, Z_{t-1}) &= \mathbb{E}(Z_t Z_{t-1}) - \underbrace{\mathbb{E} Z_t \mathbb{E} Z_{t-1}}_{=0} = \mathbb{E}((\rho Z_{t-1} + \delta_t) Z_{t-1}) \\ &= \rho \mathbb{E}(Z_{t-1}^2) + \underbrace{\mathbb{E}(\delta_t Z_{t-1})}_{\text{unabhängig}} = \rho \text{Var}(Z_{t-1}) + \underbrace{\mathbb{E} \delta_t \mathbb{E}(Z_{t-1})}_{=0} = \rho \sigma^2 . \end{aligned}$$

Somit gilt

$$\text{Cor}(Z_t, Z_{t-1}) = \frac{\text{Cov}(Z_t, Z_{t-1})}{\sqrt{\text{Var} Z_t} \sqrt{\text{Var} Z_{t-1}}} = \frac{\rho \sigma^2}{\sigma \sigma} = \rho . \quad \square$$

Im Beispiel 3.1.1 setzen wir nun voraus, dass die Störgrößen $\{\varepsilon_t\}_{t \in \mathbb{N}}$ einen AR(1)-Prozess bilden:

$$\varepsilon_t = \rho\varepsilon_{t-1} + \delta_t, \quad t \in \mathbb{Z}, \quad \delta_t \sim \mathcal{N}(0, \sigma_\delta^2) \text{ u.i.v.}, \quad |\rho| < 1.$$

Wie kann das Modell

$$Y_t = \alpha + \beta x_t + \varepsilon_t, \quad t = 1, \dots, n, \quad (3.2.1)$$

als einfache lineare Regression mit unabhängigen Störgrößen dargestellt werden?

$$\begin{cases} Y_t = \alpha + \beta x_t + \rho\varepsilon_{t-1} + \delta_t \\ Y_{t-1} = \alpha + \beta x_{t-1} + \varepsilon_{t-1} \end{cases} \Rightarrow \begin{cases} Y_t = \alpha + \beta x_t + \rho Y_{t-1} - \rho\alpha - \rho\beta x_{t-1} + \delta_t \\ \varepsilon_{t-1} = Y_{t-1} - \alpha - \beta x_{t-1} \end{cases}$$

$$\Rightarrow Y_t - \rho Y_{t-1} = \alpha(1 - \rho) + \beta(x_t - \rho x_{t-1}) + \delta_t, \quad t = 1, \dots, n,$$

Nach der Substitution
$$\begin{cases} Z_t = Y_t - \rho Y_{t-1} \\ y_{1t}^* = 1 - \rho, \quad t = 2, \dots, n, \quad Z_1 = Y_1, \quad y_{11}^* = 1, \quad y_{21}^* = x_1 \\ y_{2t}^* = x_t - \rho x_{t-1} \end{cases}$$

bekommen wir wieder ein einfaches lineares Regressionsmodell

$$Z_t = \alpha y_{1t}^* + \beta y_{2t}^* + \delta_t, \quad t = 1, \dots, n, \quad (3.2.2)$$

mit unabhängigen identisch verteilten Störgrößen δ_t .

Es können MKQ-Schätzer für α und β im Modell (3.2.1) bzw. (3.2.2) berechnet werden:

$$\hat{\alpha}^{(1)}, \hat{\beta}^{(1)} \text{ bzw. } \hat{\alpha}^{(2)}, \hat{\beta}^{(2)}.$$

Im Allgemeinen gilt $\hat{\alpha}^{(1)} \neq \hat{\alpha}^{(2)}$, $\hat{\beta}^{(1)} \neq \hat{\beta}^{(2)}$. Wegen $\mathbb{E}\varepsilon_t = \mathbb{E}\delta_t = 0 \forall t$ folgt die Unverzerrtheit (Erwartungstreue) von $\hat{\alpha}^{(i)}$, $\hat{\beta}^{(i)}$, $i = 1, 2$:

$$\mathbb{E}\hat{\alpha}^{(i)} = \alpha, \quad \mathbb{E}\hat{\beta}^{(i)} = \beta, \quad i = 1, 2.$$

Der bessere Schätzer soll die kleinste Varianz besitzen. Aus der Wirtschaftsstatistik ist bekannt, dass $\hat{\alpha}^{(2)}$, $\hat{\beta}^{(2)}$ beste lineare erwartungstreue Schätzer für α und β sind (im einfachen linearen Regressionsmodell mit unabhängigen Störgrößen). Da $\hat{\alpha}^{(1)}$, $\hat{\beta}^{(1)}$ ebenfalls linear von $\{Y_t\}$ abhängen und erwartungstreu sind, müssen sie schlechter als die besten Schätzer $\hat{\alpha}^{(2)}$, $\hat{\beta}^{(2)}$ sein:

$$\begin{aligned} \text{Var}(\hat{\alpha}^{(2)}) &< \text{Var}(\hat{\alpha}^{(1)}), \quad \text{Var}(\hat{\beta}^{(2)}) < \text{Var}(\hat{\beta}^{(1)}) \\ \Rightarrow \hat{\alpha}^{(2)}, \hat{\beta}^{(2)} &\text{ sollten an Stelle von } \hat{\alpha}^{(1)}, \hat{\beta}^{(1)} \text{ verwendet werden.} \end{aligned}$$

Praktisch aber kann das Modell (3.2.2) nicht direkt eingesetzt werden, weil ε_t und somit ρ , δ_t nicht beobachtbar, also unbekannt, sind.

$\Rightarrow \rho$ muss aus den Daten $\{Y_t\}, \{x_t\}$ geschätzt werden.

Bevor wir uns aber weiter mit der Regressionsanalyse beschäftigen, müssen wir die Zeitreihen in ihrer Allgemeinheit einführen und studieren.

3.2.2 Typen von Zeitreihen

Definition 3.2.2

Eine Folge von Zufallsvariablen $\{Z_t\}_{t \in \mathbb{Z}}$, die von Zeitparametern $t \in \mathbb{Z}$ abhängt, heißt eine *Zeitreihe*, falls es folgende Darstellung gibt:

$$Z_t = T_t + S_t + X_t, \quad t \in \mathbb{Z},$$

wobei

- T_t ein nicht zufälliges Polynom in $t \in \mathbb{R}$ ist, das *Trend* heißt.
- S_t der *saisonale Anteil* ist:

$$S_t = S_t^{(1)} + \dots + S_t^{(k)}, \quad k \geq 1,$$

wobei $S_t^{(i)} = S^{(i)}(t)$, $t \in \mathbb{R}$ periodische Funktionen sind mit Periode $p^{(i)} \in \mathbb{R}$ und der Eigenschaft

$$\int_0^{p^{(i)}} S_t^{(i)} dt = 0, \quad i = 1, \dots, k.$$

- X_t ist der *stationäre Anteil*: $\{X_t\}_{t \in \mathbb{Z}}$ ist ein zufälliger stationärer Prozess auf \mathbb{Z} . Genauer werden stationäre Prozesse etwas später eingeführt. Zur Zeit können wir behaupten, dass es solche Prozesse sind, deren stochastisches Verhalten unabhängig von den Zeitverschiebungen ist.

Beispiel 3.2.1 (Umsätze im Einzelhandel)

vgl. Löbus - Ökonometrie, S.118-129

Wir betrachten für den Zeitraum 1979 bis 1988 die Umsätze im Einzelhandel in der BRD zu den jeweiligen Preisen.

Jahr	Jan.	Feb.	März	April	Mai	Juni	Juli	Aug.	Sept.	Okt.	Nov.	Dez.
1979	69.7	66.3	83.2	79.6	81.6	80.5	75.8	75.2	75.0	86.8	90.1	104.8
1980	78.6	75.3	85.4	84.1	83.5	78.1	84.5	75.8	82.1	92.6	91.8	111.7
1981	78.8	77.9	88.8	88.5	84.7	82.5	88.5	79.2	84.3	96.5	95.2	117.6
1982	78.5	77.4	95.2	91.4	85.0	85.4	86.8	79.8	84.1	92.5	97.4	119.5
1983	79.2	78.5	101.3	88.7	90.5	93.9	84.7	83.3	91.1	96.0	101.0	121.5
1984	84.0	86.2	97.7	94.9	95.4	89.5	90.1	88.0	90.4	100.8	102.3	119.7
1985	88.1	81.7	97.4	98.4	97.5	89.4	97.8	91.3	91.2	105.2	107.5	120.8
1986	90.1	84.5	97.9	104.6	100.4	94.7	99.4	89.6	96.6	106.9	104.8	130.6
1987	90.4	87.1	101.1	109.4	100.9	98.3	106.7	94.2	99.7	112.8	113.0	136.9
1988	92.0	93.2	115.3	105.6	107.5	104.1	104.7	100.16	106.1	110.5	118.9	141.7

Tabelle 3.2.1a

In der folgenden Tabelle ist die zeitliche Entwicklung für den Preisindex der privaten Haushalte angegeben, wobei der Preisindex im Jahr 1986 = 100% ist.

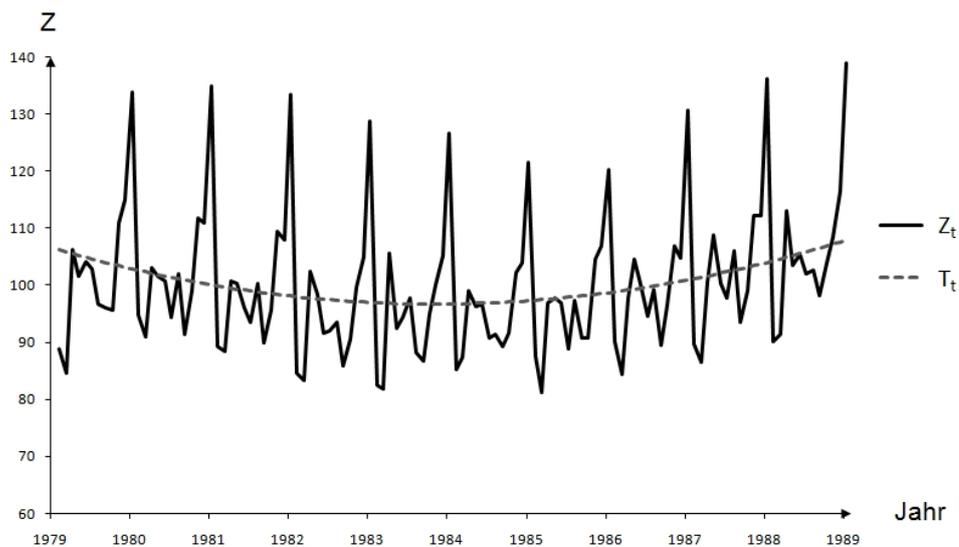
Jahr	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988
Preisindex	78.3	82.8	88.1	92.8	95.9	98.5	100.5	100.0	100.6	102.0

Tabelle 3.2.1b

Die zu analysierende Zeitreihe ist nun

$$Z_t := \frac{\text{Umsatz zur Zeit } t}{\text{Preisindex}} \cdot 100 ,$$

wobei die Zeit t in Monaten ($t = 1, \dots, 120$) gemessen wird.



Zeitreihen Z_t und ihr Trend T_t

Trend: $T_t = 0.00295t^2 - 0.342t + 106.555$, $t \in \mathbb{Z}$,

Saisonaler Anteil: $S_t = S_t^{(1)} + S_t^{(2)}$, $t \in \mathbb{Z}$,

wobei $S_t^{(1)}$ eine Funktion mit der Periode $p^{(1)} = 12$ ist, die durch folgende Tabelle gegeben ist:

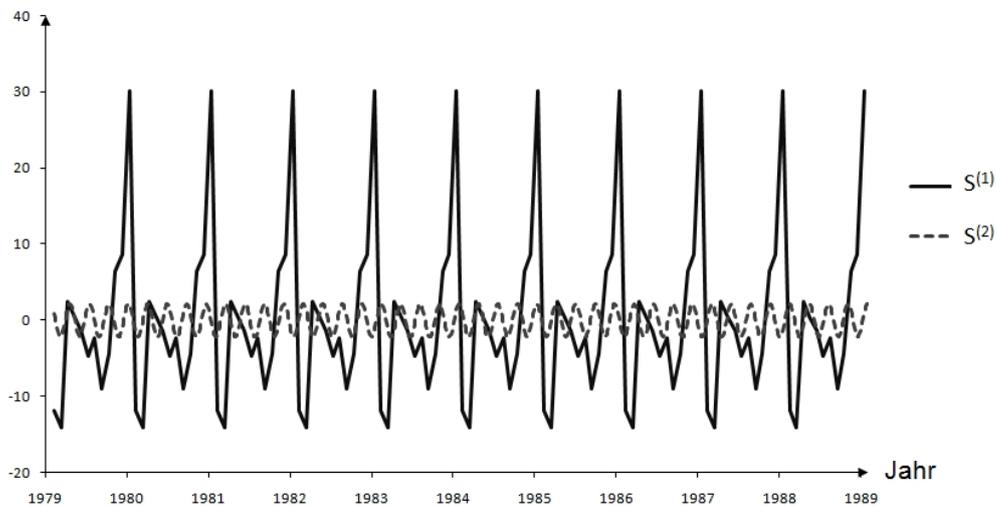
i	1	2	3	4	5	6	7	8	9	10	11	12
$S_{12k+i}^{(1)}$	-11.82	-14.12	2.42	0.41	-1.46	-4.68	-2.39	-9.05	-4.39	6.35	8.58	30.14

Tabelle 3.2.1c - Werte von $S_t^{(1)}$

Die Funktion $S^{(2)}$ ist durch

$$S_t^{(2)} = 2.14 \cos(2.20t + 1.13), \quad t \in \mathbb{R}$$

gegeben. Sie besitzt die Periode $p^{(2)} = \frac{2\pi}{2.2} = 2.86$.



Saisonale Komponente $S_t^{(1)}$ und $S_t^{(2)}$

Der stationäre Anteil der Zeitreihe Z ergibt sich als Differenz

$$X_t = Z_t - T_t - S_t, \quad t \in \mathbb{Z}.$$



Stationärer Anteil X_t von Z_t

Definition 3.2.3

Der stochastische Prozess $X = \{X_t\}_{t \in \mathbb{Z}}$ heißt *stationär* im

1. *weiteren Sinne*, falls folgende Eigenschaften gelten:

- a) $\mathbb{E} X_t^2 < \infty \quad \forall t \in \mathbb{Z}, \quad \mathbb{E} X_t = \text{const.}$
- b) $\text{Cov}(X_t, X_{s+t}) = \text{Cov}(X_0, X_s)$ hängt nicht von $t \in \mathbb{Z}$ ab.

2. *engeren Sinne*, falls alle endlich dimensionalen Verteilungen von X verschiebungsinvariant sind, d.h. $\forall n \in \mathbb{N}, \forall t_1 < t_2 < \dots < t_n \in \mathbb{Z}, \forall t \in \mathbb{Z}, \forall$ Borel-Mengen $B_1, \dots, B_n \in \mathcal{B}_{\mathbb{R}}$ gilt

$$P(X_{t+t_1} \in B_1, \dots, X_{t+t_n} \in B_n) = P(X_{t_1} \in B_1, \dots, X_{t_n} \in B_n),$$

d.h. $P(X_{t+t_1} \in B_1, \dots, X_{t+t_n} \in B_n)$ ist unabhängig von $t \in \mathbb{Z}$.

Bemerkung 3.2.2

Im Allgemeinen sind die Begriffe der Stationarität im engeren und weiteren Sinne disjunkt, d.h. aus 1) $\not\Rightarrow$ 2) und umgekehrt. So muss ein im engeren Sinne stationärer Prozess X nicht unbedingt endliche Momente 1. und 2. Ordnung besitzen.

Beispiele von stationären Prozessen

1) Lineare Prozesse

Definition 3.2.4

Sei $\{\delta_i\}_{i \in \mathbb{Z}}$ eine Folge von Zufallsvariablen mit

$$\mathbb{E} \delta_i = 0, \quad \text{Cov}(\delta_i, \delta_j) = \sigma^2 \delta_{ij} = \begin{cases} \sigma^2, & i = j \\ 0, & \text{sonst} \end{cases}. \quad (3.2.3)$$

Ferner sei $\{\gamma_i\}_{i \in \mathbb{Z}}$ eine Zahlenfolge mit $\sum_{i \in \mathbb{Z}} \gamma_i^2 < \infty$.

Definieren wir den *linearen Prozess* $\{X_t\}_{t \in \mathbb{Z}}$ als Reihe

$$X_t = \sum_{i=0}^{\infty} \gamma_i \delta_{t-i}.$$

In welchem Sinne ist $\sum_{i=0}^{\infty} \gamma_i \delta_{t-i}$ zu verstehen?

Führen wir die Zufallsvariablen

$$X_t^n = \sum_{i=0}^n \gamma_i \delta_{t-i}, \quad n \in \mathbb{N}$$

ein. Die Folge $\{X_t^n\}_{n \in \mathbb{N}}$ ist eine Cauchy-Folge in $L^2(\Omega, \mathcal{F}, P)$, wobei $L^2 = L^2(\Omega, \mathcal{F}, P)$ der *Hilbertraum* aller Zufallsvariablen auf dem Wahrscheinlichkeitsraum (Ω, \mathcal{F}, P) ist, d.h. ein vollständiger linearer normierter unendlich dimensionaler Vektorraum.

- $L^2 = \{Y - \text{Zufallsvariable auf } (\Omega, \mathcal{F}, P) : \mathbb{E} Y^2 < \infty \text{ mit Skalarprodukt } \langle Y, Z \rangle_{L^2} = \mathbb{E}(YZ)\}$.
- Norm in $L^2 : \|Y\|_{L^2} = \sqrt{\mathbb{E} Y^2}$. Wegen Vollständigkeit von $L^2(\Omega, \mathcal{F}, P)$

$$\exists X_t \stackrel{L^2}{=} \lim_{n \rightarrow \infty} X_t^n, \quad X_t \in L^2(\Omega, \mathcal{F}, P).$$

Die Konvergenz $X_t^n \xrightarrow[n \rightarrow \infty]{L^2} X_t$ bedeutet Folgendes:

$$\|X_t^n - X_t\|_{L^2} \rightarrow 0 \Leftrightarrow \mathbb{E}(X_t^n - X_t)^2 \xrightarrow[n \rightarrow \infty]{} 0,$$

d.h. die Konvergenz in L^2 ist die Konvergenz im mittleren quadratischen Sinne.

Satz 3.2.2

Sei $X = \{X_t\}_{t \in \mathbb{Z}}$ ein linearer Prozess:

$$X_t = \sum_{i=0}^{\infty} \gamma_i \delta_{t-i}.$$

Dann ist X stationär im weiteren Sinne, d.h.

$$\mathbb{E} X_t = 0, \quad \forall t \in \mathbb{Z}, \quad \text{Cov}(X_t, X_{t+s}) = \mathbb{E}(X_t X_{t+s}) = \sigma^2 \sum_{i=0}^{\infty} \gamma_i \gamma_{i+|s|}$$

Beweis

1. Zeigen wir, dass $\mathbb{E} X_t = 0, \forall t \in \mathbb{Z}$. Wir wissen, dass $X_t^n \xrightarrow[n \rightarrow \infty]{L^2} X_t$, wobei

$$\mathbb{E} X_t^n = \mathbb{E} \left(\sum_{i=0}^n \gamma_i \delta_{t-i} \right) = \sum_{i=0}^n \gamma_i \underbrace{\mathbb{E} \delta_{t-i}}_{=0} = 0, \quad \text{somit gilt}$$

$$\mathbb{E} X_t = \mathbb{E} X_t - \underbrace{\mathbb{E} X_t^n}_{=0} = \mathbb{E}(X_t - X_t^n) = \underbrace{\langle X_t - X_t^n, 1 \rangle_{L^2}}_{\xrightarrow[n \rightarrow \infty]{L^2} 0} \xrightarrow[n \rightarrow \infty]{} 0, \quad \text{weil}$$

für $\{a_n\} : a_n \rightarrow 0$ folgt $(a_n, b) \rightarrow 0$. Deshalb gilt $\mathbb{E} X_t \equiv 0, \forall t \in \mathbb{Z}$.

2. Zeigen wir, dass $\text{Cov}(X_t, X_{t+s}) = \sigma^2 \sum_{i=0}^{\infty} \gamma_i \gamma_{i+|s|}, \forall s \in \mathbb{Z}$.

$$\begin{aligned} \text{Cov}(X_t^n, X_{t+s}^n) &= \mathbb{E}(X_t^n X_{t+s}^n) = \mathbb{E} \left(\sum_{i=0}^n \gamma_i \delta_{t-i} \sum_{j=0}^n \gamma_j \delta_{t+s-j} \right) \\ &= \mathbb{E} \left(\sum_{i,j=0}^n \gamma_i \gamma_j \delta_{t-i} \delta_{t+s-j} \right) = \sum_{i,j=0}^n \gamma_i \gamma_j \mathbb{E}(\delta_{t-i} \delta_{t+s-j}) \\ &= \sum_{i=0}^n \gamma_i \gamma_{i+s} \sigma^2 \xrightarrow[n \rightarrow \infty]{} \sigma^2 \sum_{i=0}^{\infty} \gamma_i \gamma_{i+s} = \sigma^2 \sum_{i=0}^{\infty} \gamma_i \gamma_{i+|s|}. \end{aligned}$$

Zeigen wir schließlich, dass $\text{Cov}(X_t^n, X_{t+s}^n) \xrightarrow[n \rightarrow \infty]{} \text{Cov}(X_t, X_{t+s}), \forall t, s \in \mathbb{Z}$.

$$\begin{aligned} |\mathbb{E}(X_t X_{t+s}) - \mathbb{E}(X_t^n X_{t+s}^n)| &= |\mathbb{E}(X_t X_{t+s}) - \mathbb{E}(X_t X_{t+s}^n) + \mathbb{E}(X_t X_{t+s}^n) - \mathbb{E}(X_t^n X_{t+s}^n)| \\ &= |\mathbb{E}(X_t(X_{t+s} - X_{t+s}^n)) + \mathbb{E}(X_{t+s}^n(X_t - X_t^n))| \\ &= |(\underbrace{X_t, X_{t+s} - X_{t+s}^n}_{\xrightarrow[n \rightarrow \infty]{L^2} 0})_{L^2} + (\underbrace{X_{t+s}^n, X_t - X_t^n}_{\xrightarrow[n \rightarrow \infty]{L^2} 0})_{L^2}| \\ &\stackrel{4}{\leq} \|X_t\|_{L^2} \cdot \|X_{t+s} - X_{t+s}^n\|_{L^2} + \underbrace{\|X_{t+s}^n\|_{L^2}}_{\|X_{t+s}\|} \cdot \underbrace{\|X_t - X_t^n\|}_{\xrightarrow[n \rightarrow \infty]{L^2} 0} \xrightarrow[n \rightarrow \infty]{} 0 \\ &\Rightarrow \text{Cov}(X_t, X_{t+s}) = \sigma^2 \sum_{i=0}^{\infty} \gamma_i \gamma_{i+|s|} \quad \square \end{aligned}$$

⁴Ungleichung von Cauchy-Schwarz

2) Prozesse der gleitenden Mittel

Definition 3.2.5

Sei $\alpha_0, \dots, \alpha_q$ eine endliche Zahlenfolge mit $\alpha_0 = 1$, $\alpha_q \neq 0$, $q \in \{0, 1, 2, \dots\}$. Sei $\{\delta_t\}_{t \in \mathbb{Z}}$ eine Folge von Zufallsvariablen mit den Eigenschaften (3.2.3). Der stochastische Prozess

$$X = \{X_t\}_{t \in \mathbb{Z}} : X_t = \sum_{i=0}^q \alpha_i \delta_{t-i}, \quad t \in \mathbb{Z}$$

heißt *Prozess der gleitenden Mittel* (engl. *moving average process*, MA(q)) der Ordnung q . Es folgt aus dem Satz 3.2.2, dass MA(q)-Prozesse stationär (im weiteren Sinne) sind. Es ist klar, dass $X_t \in L^2(\Omega, \mathcal{F}, P)$.

3) Autoregressive Prozesse

Im Abschnitt 3.1 haben wir einen stationären AR(1)-Prozess betrachtet. Im Folgenden soll eine allgemeine Definition eines nicht-stationären AR(p)-Prozesses gegeben werden.

Definition 3.2.6

Sei β_0, \dots, β_p eine endliche Zahlenfolge mit $\beta_0 = 1$, $\beta_p \neq 0$ für ein $p \in \{0, 1, 2, \dots\}$. Der stochastische Prozess $X = \{X_t\}_{t \in \mathbb{Z}}$ heißt *autoregressiver Prozess der Ordnung p* (AR(p)), falls X eine Lösung folgender Differenzgleichung ist:

$$\sum_{i=0}^p \beta_i X_{t-i} = \delta_t, \quad t \in \mathbb{Z},$$

wobei $\{\delta_t\}_{t \in \mathbb{Z}}$ eine Folge von Zufallsvariablen mit den Eigenschaften (3.2.3) ist.

Nicht jeder AR(p)-Prozess ist stationär. Die stationären AR(p)-Prozesse werden jedoch im folgenden Satz charakterisiert:

Satz 3.2.3

Sei $X = \{X_t\}_{t \in \mathbb{Z}}$ ein oben eingeführter AR(p)-Prozess mit folgender Eigenschaft: alle Wurzeln $z_1, \dots, z_p \in \mathbb{C}$ des Polynoms

$$P(z) = \sum_{i=0}^p \beta_i z^{p-i}, \quad z \in \mathbb{C},$$

liegen im Einheitskreis $B_1(0) = \{z \in \mathbb{C} : |z| \leq 1\}$, d.h. $z_j \in B_1(0) \forall j = 1, \dots, p$.

a) Dann existiert eine Folge

$$\{\gamma_i\}_{i=0}^{\infty} \text{ mit } \gamma_0 = 1, \quad \sum_{i=0}^{\infty} \gamma_i^2 < \infty,$$

so dass

$$X_t = \sum_{i=0}^{\infty} \gamma_i \delta_{t-i}$$

ein linearer Prozess ist. Nach dem Satz 3.2.2 ist X stationär.

b) Dabei gilt

$$\gamma_i = \sum_{\substack{k_1, \dots, k_p \in \mathbb{N} \cup \{0\}: \\ k_1 + \dots + k_p = i}}^n z_1^{k_1} \cdot \dots \cdot z_p^{k_p} \in \mathbb{R}, \quad i \in \mathbb{N}, \quad |\gamma_i| \leq c \left(\frac{1 + \zeta}{2} \right)^i, \quad i \in \mathbb{N}$$

für eine Konstante $c > 0$ und $\zeta = \max_{j=1, \dots, p} |z_j|$.

Ein Beweis befindet sich in „Löbus - Ökonometrie“.

4) Autoregressive Prozesse mit gleitendem Mittel

Definition 3.2.7

Für $p, q \in \{0, 1, 2, \dots\}$ seien $\alpha_0, \dots, \alpha_q$ bzw. β_0, \dots, β_p Zahlenfolgen mit $\alpha_q \neq 0, \beta_p \neq 0$. Der stochastische Prozess $X = \{X_t\}_{t \in \mathbb{Z}}$ heißt *autoregressiver Prozess mit gleitendem Mittel* (ARMA(p, q)), falls er eine Lösung der Differenzgleichung

$$\sum_{i=0}^p \beta_i X_{t-i} = \sum_{j=0}^q \alpha_j \delta_{t-j}, \quad \forall t \in \mathbb{Z},$$

darstellt.

Man kann zeigen, dass der Satz 3.2.3 auch für ARMA(p, q)-Prozesse gilt.

5) Integrierte autoregressive Prozesse mit gleitendem Mittel

Definition 3.2.8

Seien $p, q \in \{0, 1, 2, \dots\}$. Der stationäre Prozess $X = \{X_t\}_{t \in \mathbb{Z}}$ heißt *integrierter autoregressiver Prozess mit gleitendem Mittel* (engl. *autoregressive integrated moving average process*, ARIMA(p, q)), falls der Prozess $\Delta X = \{\Delta X_t\}_{t \in \mathbb{Z}}$ der Zuwächse

$$\Delta X_t = X_t - X_{t-1}, \quad t \in \mathbb{Z} \quad \text{von } X$$

ein ARMA(p, q) - Prozess ist. Allgemeiner, für den rekurrent definierbaren Operator

$$\Delta^r, \quad \text{für den } \Delta^0 = \text{Id}, \Delta^r = \Delta(\Delta^{r-1}), \quad r \in \mathbb{N} \cup \{0\} \text{ gilt:}$$

$X = \{X_t\}_{t \in \mathbb{Z}}$ ist ein ARIMA(p, q, r)-Prozess, falls $\Delta^r X = \{\Delta^r X_t\}_{t \in \mathbb{Z}}$ ein ARMA(p, q)-Prozess ist.

3.2.3 Vorhersage von Zeitreihen

Hier wird die stationäre Komponente $X = \{X_t\}_{t \in \mathbb{Z}}$ einer Zeitreihe $Z = \{Z_t\}_{t \in \mathbb{Z}}$ betrachtet. Im Folgenden wird der *Zerlegungssatz von Wold* formuliert, der die besondere Stellung der linearen Prozesse in der Klasse aller stationären Prozesse unterstreicht. Zuvor aber werden wir einige Begriffe einführen müssen, die die Vorhersage von stationären Prozessen betreffen. Uns wird vor allem interessieren, ob der Wert X_t des Prozesses X eindeutig aus der Vergangenheit X_{t-1}, \dots, X_{t-p} vorhersagbar ist oder dessen Vorhersage Unsicherheiten enthält.

Definition 3.2.9

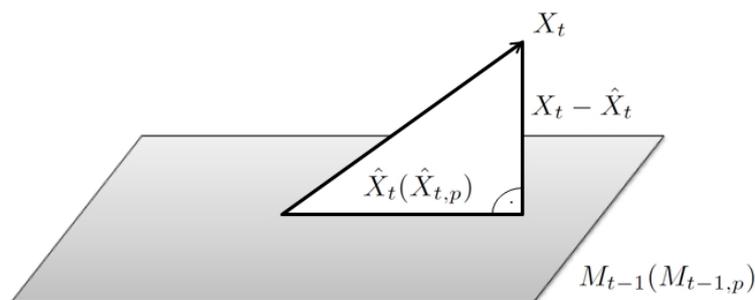
Sei $X = \{X_t\}_{t \in \mathbb{Z}}$ ein stationärer Prozess in $L^2(\Omega, \mathcal{F}, P)$.

1. Der *Unterraum* M_{t-1} , $t \in \mathbb{Z}$, wird definiert als linearer Abschluss in $L^2(\Omega, \mathcal{F}, P)$ von der Folge $\{X_{t-1}, X_{t-2}, \dots\}$, die die Vergangenheit von X bis zum Zeitpunkt $t \in \mathbb{Z}$ darstellt, d.h. M_{t-1} enthält alle Linearkombinationen der Zufallsvariablen aus der Folge $\{X_{t-1}, X_{t-2}, \dots\}$ und alle möglichen Grenzwerte von Folgen von solchen Zufallsvariablen.
2. Der *Unterraum* $M_{t-1,p}$, $t \in \mathbb{Z}$, $p \in \mathbb{N}$, ist die lineare Hülle $\langle X_{t-1}, \dots, X_{t-p} \rangle$ von Zufallsvariablen X_{t-1}, \dots, X_{t-p} , d.h. $M_{t-1,p}$ besteht aus allen Linearkombinationen von X_{t-1}, \dots, X_{t-p} . $M_{t-1,p}$ stellt somit die Vergangenheit von X bis zum Zeitpunkt $t - p$ dar.

Diese Räume werden eingeführt, weil wir uns für die lineare Prognose des Wertes X_t aus der Vergangenheit M_{t-1} bzw. $M_{t-1,p}$ interessieren:

$$\sum_{i=0}^{\infty} \lambda_i X_{t-i}, \quad \text{für } \lambda_i \in \mathbb{R}, \quad i \in \mathbb{N}.$$

Anschaulich sucht man die Vorhersage von X_t als die orthogonale Projektion von X_t auf M_{t-1} bzw. $M_{t-1,p}$.



Definition 3.2.10

1. Die Orthogonalprojektion $\hat{X}_t = Pr_{M_{t-1}}(X_t)$ von X_t auf M_{t-1} heißt *Vorhersage* von X_t durch die Vergangenheit X_{t-1}, X_{t-2}, \dots von X mit unendlichem Zeithorizont:

$$\hat{X}_t = \operatorname{argmin}_{Y \in M_{t-1}} \|X_t - Y\|_{L^2} = \operatorname{argmin}_{Y \in M_{t-1}} \mathbb{E} |X_t - Y|^2 .$$

2. Die Orthogonalprojektion $\hat{X}_{t,p} = Pr_{M_{t-1,p}}(X_t)$ von X_t auf $M_{t-1,p}$ heißt *Vorhersage* von X_t durch die Vergangenheit X_{t-1}, \dots, X_{t-p} von X mit endlichem Zeithorizont und $p \in \mathbb{N}$:

$$\hat{X}_{t,p} = \operatorname{argmin}_{Y \in M_{t-1,p}} \|X_t - Y\|_{L^2} = \operatorname{argmin}_{Y \in M_{t-1,p}} \mathbb{E} |X_t - Y|^2 .$$

Jetzt untersuchen wir die Güte dieser Prognose, die sich durch

$$\mathbb{E} |X_t - \hat{X}_t|^2 \text{ bzw. } \mathbb{E} |X_t - \hat{X}_{t,p}|^2 \text{ und } \operatorname{Var}(\hat{X}_t) \text{ bzw. } \operatorname{Var}(\hat{X}_{t,p})$$

quantifizieren lässt. Je kleiner diese Größen sind, desto besser ist die Prognose.

Man kann Folgendes zeigen:

Lemma 3.2.1

Sei $X = \{X_t\}_{t \in \mathbb{Z}}$ ein stationärer Prozess⁵ aus $L^2(\Omega, \mathcal{F}, P)$ mit $\mathbb{E} X_t \equiv 0$. Dann hängt $\mathbb{E} (X_t - \hat{X}_t)^2$ nicht von t ab.

(ohne Beweis)

Deshalb genügt es, $\mathbb{E} (X_0 - \hat{X}_0)^2$ zu betrachten, um die Güte der Prognose \hat{X}_t zu beurteilen.

Definition 3.2.11

Sei $X = \{X_t\}_{t \in \mathbb{Z}}$ ein stationärer Prozess aus $L^2(\Omega, \mathcal{F}, P)$ mit $\mathbb{E} X_t \equiv 0$.

1. X heißt *regulär*, falls $\mathbb{E} (X_0 - \hat{X}_0)^2 > 0$, d.h. X lässt sich nicht ohne intrinsische Unsicherheit in die Zukunft vorhersagen.
2. X heißt *singulär*, falls $\mathbb{E} (X_0 - \hat{X}_0)^2 = 0$, d.h. es ist immer eine genaue Prognose $\hat{X}_t \stackrel{\text{f.s.}}{=} X_t$ (im $L^2(\Omega, \mathcal{F}, P)$ -Sinne) aus der Vergangenheit möglich:

$$P(\hat{X}_t = X_t) = 1 .$$

Lemma 3.2.2

Für einen stationären Prozess $X = \{X_t\}_{t \in \mathbb{Z}}$ aus $L^2(\Omega, \mathcal{F}, P)$ mit $\mathbb{E} X_t \equiv 0$ gilt

$$\lim_{p \rightarrow \infty} \mathbb{E} |\hat{X}_t - \hat{X}_{t,p}|^2 = 0 ,$$

⁵es handelt sich im Folgenden immer um einen stationärer Prozess im weiteren Sinne

d.h. die Prognose X mit endlichem Zeithorizont p konvergiert gegen die Prognose \hat{X}_t mit unendlichem Zeithorizont, wenn die Anzahl p der beobachteten Größen X_{t-s} unendlich steigt.

Satz 3.2.4 (Woldscher Zerlegungssatz)

Sei $X = \{X_t\}_{t \in \mathbb{Z}}$ ein stationärer Prozess aus $L^2(\Omega, \mathcal{F}, P)$ mit $\mathbb{E} X_t \equiv 0$. Dann gilt folgende Darstellung:

$$X_t = R_t + S_t, \quad t \in \mathbb{Z},$$

wobei

$$R_t = \sum_{i=0}^{\infty} \gamma_i \delta_{t-i}, \quad t \in \mathbb{Z}$$

ein regulärer linearer Prozess ist, mit

$$\{\gamma_i\}_{i \in \mathbb{N}} \subset \mathbb{R}, \quad \sum_{i=1}^{\infty} \gamma_i^2 < \infty, \quad \gamma_0 = 1,$$

einer Folge von Zufallsvariablen

$$\{\delta_i\}_{i \in \mathbb{Z}} : \mathbb{E} \delta_i = 0, \quad \mathbb{E} (\delta_i \delta_j) = \begin{cases} \sigma^2, & i = j \\ 0, & i \neq j \end{cases},$$

$$\{\alpha \delta_t : \alpha \in \mathbb{R}\} = M_t \cap M_{t-1}^{\perp},$$

$$M_{t-1}^{\perp} = \{\text{Zufallsvariable } Y \in L^2(\Omega, \mathcal{F}, P) : \mathbb{E}(ZY) = 0 \forall Z \in M_{t-1}\}$$

und $S = \{S_t\}_{t \in \mathbb{Z}}$, ein singulärer stationärer Prozess ist, mit der Eigenschaft

$$\mathbb{E} S_t = 0 \quad \forall t \in \mathbb{Z},$$

$$\mathbb{E} (\delta_s S_t) = 0 \quad \forall s, t \in \mathbb{Z} \quad (\{\delta_i\}_{i \in \mathbb{Z}} \text{ und } S \text{ sind unkorreliert}),$$

$$S_t \in \bigcap_{s=0}^{\infty} M_{t-s} = M_{-\infty}, \quad t \in \mathbb{Z}$$

$R = \{R_t\}_{t \in \mathbb{Z}}$ heißt dabei *rein regulärer* Anteil von X .

Wie funktioniert die lineare Prognose $\hat{X}_{t,p}$ genau?

Satz 3.2.5

1. Sei $X = \{X_t\}_{t \in \mathbb{Z}}$ ein stationärer Prozess aus $L^2(\Omega, \mathcal{F}, P)$ mit $\mathbb{E} X_t \equiv 0$. Für $p \in \mathbb{N}$ sei X_{t_1}, \dots, X_{t_q} , ($q \leq p$), eine Basis von $M_{-1,p} = \langle X_{-1}, \dots, X_{-p} \rangle$.

Sei $K(s) = \mathbb{E}(X_0 X_s)$ die Kovarianzfunktion des Prozesses X . Es gilt

$$\hat{X}_{t,p} = (X_{t+t_1}, \dots, X_{t+t_q}) \cdot (K_q)^{-1} \cdot (K(t_1), \dots, K(t_q))^{\top},$$

wobei

$$K_q = (K(t_i - t_j))_{i,j=1,\dots,q} = \begin{pmatrix} K(0) & K(t_1 - t_2) & \dots & K(t_1 - t_q) \\ \vdots & \vdots & \ddots & \vdots \\ K(t_q - t_1) & K(t_q - t_2) & \dots & K(0) \end{pmatrix}.$$

2. Sei $S = \{S_t\}_{t \in \mathbb{Z}}$ ein singulärer Prozess aus $L^2(\Omega, \mathcal{F}, P)$ mit $\mathbb{E} S_t \equiv 0 \forall t \in \mathbb{Z}$.

$K^S(t) = \mathbb{E}(S_0 S_t)$, $t \in \mathbb{Z}$. Sei S_{t_1}, \dots, S_{t_q} eine Basis von $\langle S_{-1}, \dots, S_{-p} \rangle$, $q \leq p$. Dann gilt

$$S_t = \lim_{p \rightarrow \infty} \hat{S}_{t,p} = \lim_{p \rightarrow \infty} (S_{t+t_1}, \dots, S_{t+t_q}) \cdot (K_q^S)^{-1} \cdot (K^S(t_1), \dots, K^S(t_q))^T,$$

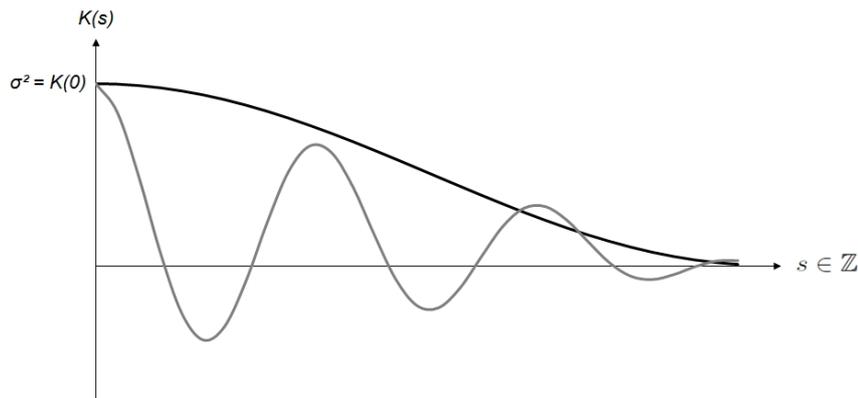
wobei

$$K_q^S = (K_q^S(t_i - t_j))_{i,j=1,\dots,q} = \begin{pmatrix} K^S(0) & K^S(t_1 - t_2) & \dots & K^S(t_1 - t_q) \\ \vdots & \vdots & \ddots & \vdots \\ K^S(t_q - t_1) & K^S(t_q - t_2) & \dots & K^S(0) \end{pmatrix}.$$

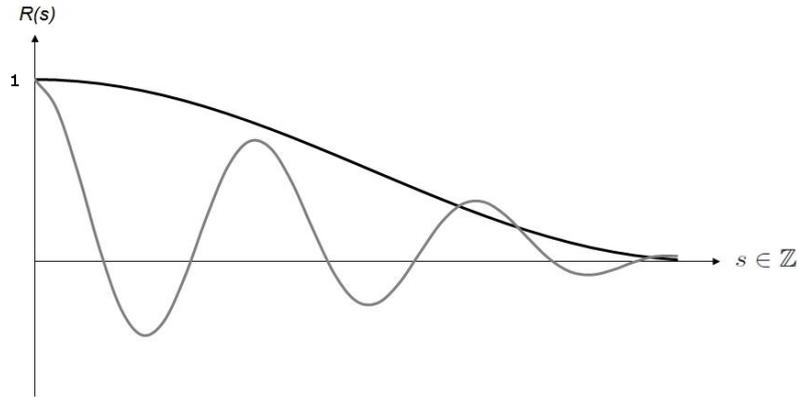
3.2.4 Eigenschaften der Zeitreihen

Definition 3.2.12

Sei $X = \{X_t\}_{t \in \mathbb{Z}}$ ein stationärer Prozess mit $\mathbb{E} X_t \equiv 0$ und $\sigma^2 = \text{Var} X(t) > 0$. Die Funktion $K(s) = \mathbb{E}(X_0 X_s)$ bzw. $R(s) = K(s)/\sigma^2$ heißt (Auto-)Kovarianz- bzw. (Auto-)Korrelationsfunktion von X .



(Auto-)Kovarianzfunktion



(Auto-)Korrelationsfunktion

Es gilt: $K(s) = K(-s)$, $R(s) = R(-s)$.

Beweis

$$K(-s) = \mathbb{E}(X_0 X_{-s}) \stackrel{X \text{ stationär}}{=} \mathbb{E}(X_{0+s} X_{s-s}) = \mathbb{E}(X_s X_0) = K(s), \quad \forall s \in \mathbb{Z}.$$

Deshalb genügt es, die Grafik von $K(s)$ bzw. $R(s)$ für $s \geq 0$ zu zeichnen.

Beispiele

1) Lineare Prozesse

$$X_t = \sum_{i=0}^{\infty} \gamma_i \delta_{t-i}, \quad \sigma^2 = \text{Var } \delta_j.$$

Aus dem Satz 3.2.2 folgt

$$K(s) = \sigma^2 \sum_{j=0}^{\infty} \gamma_{j+|s|} \gamma_j, \quad R(s) = \frac{\sum_{j=0}^{\infty} \gamma_{j+|s|} \gamma_j}{\sum_{j=0}^{\infty} \gamma_j^2}.$$

Zeigen wir, dass $K(s) \xrightarrow{s \rightarrow +\infty} 0$.

$$|K(s)| \stackrel{6}{\leq} \sigma^2 \sqrt{\sum_{j=0}^{\infty} \gamma_{j+|s|}^2 \sum_{j=0}^{\infty} \gamma_j^2} = \sigma^2 \sqrt{\sum_{i=|s|}^{\infty} \gamma_i^2 \sum_{j=0}^{\infty} \gamma_j^2} \xrightarrow{s \rightarrow +\infty} 0,$$

weil

$$\sum_{i=|s|}^{\infty} \gamma_i^2 \xrightarrow{|s| \rightarrow \infty} 0 \quad \text{wegen} \quad \sum_{i=0}^{\infty} \gamma_i^2 < \infty.$$

⁶Ungleichung von Cauchy-Schwarz

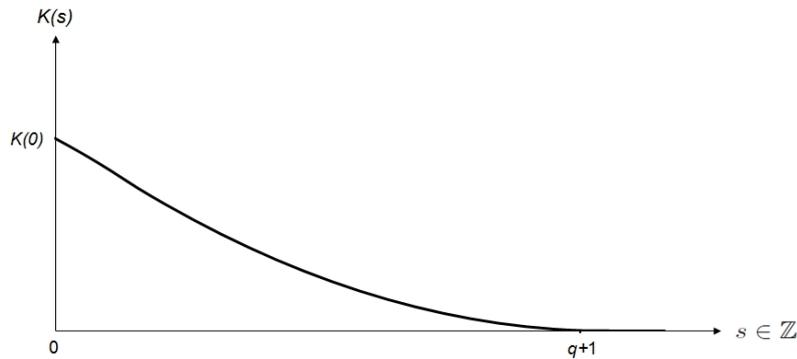
2) MA(q)-Prozesse

$$X_t = \sum_{i=0}^q \alpha_i \delta_{t-i}$$

$$\Rightarrow K(s) = \begin{cases} 0, & |s| > q \\ \sigma^2 \sum_{j=0}^{q-|s|} \alpha_{j+|s|} \alpha_j, & |s| \leq q \end{cases},$$

was aus 1) für $\gamma_{q+1} = \gamma_{q+2} = \dots = 0$ folgt.

$$R(s) = \begin{cases} 0, & |s| > q \\ \sum_{j=0}^{q-|s|} \alpha_{j+|s|} \alpha_j / \sum_{j=0}^q \alpha_j^2, & |s| \leq q \end{cases}.$$



Satz 3.2.6

Sei $X = \{X_t\}_{t \in \mathbb{Z}}$ ein stationärer Prozess aus $L^2(\Omega, \mathcal{F}, P)$ mit $\mathbb{E} X_t \equiv 0$, $K(0) = \text{Var } X_t > 0$ mit (Auto-)Kovarianz $K(s)$, $s \in \mathbb{Z}$. X ist ein MA(q)-Prozess genau dann, wenn $K(q) \neq 0$, $K(s) = 0$, $|s| > q$. In diesem Fall ist X regulär und es gelten folgende Eigenschaften:

$$X_t = \sum_{i=0}^q \alpha_i \delta_{t-i}, \quad t \in \mathbb{Z}, \quad \text{wobei } \delta_t = X_t - \hat{X}_t, \quad t \in \mathbb{Z}, \quad \sigma^2 = \text{Var } \delta_t,$$

$$\alpha_i = \frac{\mathbb{E}(X_t \delta_{t-i})}{\sigma^2}, \quad i \in \{0, \dots, q\}, \quad \mathbb{E} \delta_t = 0, \quad \mathbb{E}(\delta_i \delta_j) = \sigma^2 \delta_{ij}, \quad \forall i, j \in \mathbb{Z}.$$

(ohne Beweis)

3) Stationäre AR(p)-Prozesse

Sie werden definiert als Lösung des Gleichungssystems

$$\sum_{j=0}^p \beta_j X_{t-j} = \delta_t, \quad t \in \mathbb{Z}. \quad X\text{-stationär} \Rightarrow \exists \{\gamma_j\}_{j=0}^{\infty} \text{ mit } \sum_{j=0}^{\infty} \gamma_j^2 < \infty,$$

so dass

$$X_t = \sum_{i=0}^{\infty} \gamma_j \delta_{t-j} \Rightarrow \text{Nach 1) kann } K(s), R(s) \text{ berechnet werden.}$$

Allerdings sind γ_j nicht explizit als Funktion von $\{\beta_0, \dots, \beta_p\}$ gegeben. Im Folgenden geben wir ein Verfahren an, das die Werte $K(s)$ bzw. $R(s)$ als Funktion von $\{\beta_0, \dots, \beta_p\}$ liefert.

Satz 3.2.7 (Yule-Walker-Gleichungen)

Sei $\beta_0 = 1$. Die Werte $R(1), \dots, R(p-1)$ sind die Lösung des folgenden linearen Gleichungssystems:

$$\begin{pmatrix} \beta_0 + \beta_2 & \beta_3 & \beta_4 & \dots & \beta_{p-1} & \beta_p \\ \beta_1 + \beta_3 & \beta_0 + \beta_4 & \beta_5 & \dots & \beta_p & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \beta_{p-2} + \beta_p & \beta_{p-3} & \beta_{p-4} & \dots & \beta_1 & \beta_0 \end{pmatrix} \cdot \begin{pmatrix} R(1) \\ R(2) \\ \vdots \\ R(p-1) \end{pmatrix} = - \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{p-1} \end{pmatrix} \quad (3.2.4)$$

Zusätzlich gilt:

$$K(0) = \frac{\sigma^2}{\sum_{j=0}^p \beta_j R(j)}, \quad \text{wobei } \sigma^2 = \text{Var } \delta_j.$$

Weitere Werte $R(t)$, $t \geq p$ bekommt man aus den Werten $R(t)$, $t < p$ durch folgende Gleichung:

$$\sum_{j=0}^p \beta_j R(t-j) = 0 \quad (\text{Yule-Walker-Gleichung})$$

z.B.

$$R(p) = - \sum_{j=1}^p \beta_j R(p-j), \quad \text{weil } \beta_0 = 1. \quad (3.2.5)$$

Beweis

Aus der Definition von $\text{AR}(p)$ gilt

$$\delta_t = \beta_0 X_t + \dots + \beta_p X_{t-p}$$

Aus der linearen Darstellung von X

$$\sum_{j=0}^{\infty} \gamma_j \delta_{-j} = X_0$$

Durch Ausmultiplizieren der beiden Seiten und Berechnung des Erwartungswertes gilt

$$0 = \sum_{j=0}^{\infty} \underbrace{\mathbb{E}(\delta_t \delta_{-j})}_{=0} \gamma_j = \beta_0 \underbrace{\mathbb{E}(X_0 X_t)}_{K(t)} + \dots + \beta_p \underbrace{\mathbb{E}(X_0 X_{t-p})}_{=K(t-p)}, \quad t > 0. \quad (3.2.6)$$

wegen der
Unkorreliertheit
von $\{\delta_t\}$

Durch die Division mit $K(s)$ erhalten wir die Yule-Walker-Gleichung:

$$\sum_{j=0}^{t-p} \beta_j R(t-j) = 0 .$$

Für $t = p+1, \dots, 2p-1$

$$\left. \begin{array}{l} \sum_{j=0}^l \beta_j R(p+1-j) = 0 \\ \vdots \\ \sum_{j=0}^{p-1} \beta_j R(2p-1-j) = 0 \end{array} \right| \Rightarrow \text{Ergebnis (3.2.4).}$$

Um $K(0)$ zu berechnen, verwenden wir den Vorgang (3.2.6) für $t = 0$:

$$\underbrace{\gamma_0}_{=1} \underbrace{\mathbb{E} \delta_0^2}_{=\sigma^2} = \beta_0 K(0) + \dots + \beta_p K(-p)$$

$$\frac{\sigma^2}{K(0)} = \sum_{j=0}^p \beta_j R(j) \Rightarrow K(0) = \frac{\sigma^2}{\sum_{j=0}^p \beta_j R(j)}$$

□

4) Stationäre ARMA(p, q)-Prozesse

Der Ausdruck für $K(s)$ bzw. $R(s)$ als Funktion von $\{\alpha_0, \dots, \alpha_q\}$ und $\{\beta_0, \dots, \beta_p\}$ folgt aus der Darstellung von X als linearer Prozess mit Hilfe des Satzes 3.2.2.

Für weitere Details siehe *Löbus - Ökonometrie*

Satz 3.2.8 (Bochner)

Sei $K(s), s \in \mathbb{Z}$, eine Kovarianzfunktion eines stationären Prozesses X . Dann existiert eine nichtfallende linksseitig stetige Funktion $F : [-\pi; \pi] \rightarrow [0; K(0)]$ mit den Eigenschaften $F(-\pi) = 0, F(\pi) = K(0) = \text{Var } X_t$, so dass

$$K(s) = \int_{-\pi}^{\pi} e^{its} dF(t), \quad s \in \mathbb{Z}. \quad (\text{Spektraldarstellung von } K)$$

Definition 3.2.13

1. Die Funktion F aus der Spektraldarstellung von K heißt *Spektralfunktion* von X .
2. Falls $dF(t)$ absolut stetig bzgl. des Lebesguemaßes ist, d.h.

$$K(s) = \int_{-\pi}^{\pi} e^{its} f(t) dt, \quad s \in \mathbb{Z},$$

dann heißt f die *Spektraldichte* von X . $\Rightarrow f$ ist eine symmetrische Funktion:

$$f(-s) = f(s), \quad s \in \mathbb{Z}.$$

Beispiele

1. Lineare Prozesse

$$X_t = \sum_{j=0}^{\infty} \gamma_j \delta_{t-j}, \quad t \in \mathbb{Z}, \quad \sum_{j \in \mathbb{Z}} \gamma_j^2 < \infty$$

$\Rightarrow X$ besitzt eine Spektraldichte

$$f(s) = \frac{\sigma^2}{2\pi} \left| \sum_{j=0}^{\infty} \gamma_j e^{-isj} \right|^2$$

2. MA(q)-Prozesse

$$X_t = \sum_{i=0}^q \alpha_i \delta_{t-i}, \quad t \in \mathbb{Z}, \quad \alpha_0 = 1$$

$\Rightarrow X$ besitzt eine Spektraldichte

$$f(s) = \frac{\sigma^2}{2\pi} \left| \sum_{j=0}^q \alpha_j e^{-isj} \right|^2$$

3. AR(p)-Prozesse

Falls $X = \{X_t\}_{t \in \mathbb{Z}}$ ein stationärer AR(p)-Prozess mit Kovarianzfunktion $K(s)$ ist, dann besitzt $K(s)$ folgende Spektraldichte:

$$f(s) = \frac{\sigma^2}{2\pi} \left| \sum_{j=0}^p \beta_j e^{-isj} \right|^{-2},$$

wobei $\sigma^2 = \text{Var } \delta_j$, $\{\beta_0, \dots, \beta_p\}$ die Koeffizienten aus der Definition 3.2.6 von X sind.

4. ARMA(p, q)-Prozesse

Falls $X = \{X_t\}_{t \in \mathbb{Z}}$ ein stationärer ARMA(p, q)-Prozess ist mit Kovarianzfunktion $K(s)$, dann besitzt $K(s)$ die Spektraldichte:

$$f(s) = \frac{\sigma^2}{2\pi} \frac{\left| \sum_{j=0}^q \alpha_j e^{-isj} \right|^2}{\left| \sum_{j=0}^p \beta_j e^{-isj} \right|^2},$$

wobei $\sigma^2 = \text{Var } \delta_j$, $\{\alpha_0, \dots, \alpha_q\}$ und $\{\beta_0, \dots, \beta_p\}$ die Koeffizienten aus der Definition 3.2.7 von X sind.

Zur Erinnerung:

$$K(s) = \int_{-\pi}^{\pi} e^{its} f(t) dt .$$

Ein weiterer möglicher Ansatz zur Schätzung von $K(s)$ (abgesehen von (3.1.1)) ist der *spektrale Ansatz*, bei dem $K(s)$ als

$$\hat{K}(s) = \int_{-\pi}^{\pi} e^{its} \hat{f}(t) dt$$

geschätzt wird, wobei $\hat{f}(t)$ eine Schätzung der Spektraldichte darstellt.

3.3 Statistik in der Zeitreihenanalyse

Wir beschäftigen uns zunächst mit Statistik der stationären Komponenten von Zeitreihen. Da aber stationäre Prozesse im Wesentlichen linear sind (vgl. den Satz 3.2.4 bzw. 3.2.5), genügt es, lineare Prozesse zu betrachten. Da aber ein beliebiger linearer Prozess

$$X_t = \sum_{j=0}^{\infty} \gamma_j \delta_{t-j}, \quad t \in \mathbb{Z}$$

eine Reihendarstellung mit unendlich vielen unbekanntem Koeffizienten γ_j besitzt, müssen engere Klassen von Prozessen (wie z.B. $MA(q)$, $AR(p)$, $ARMA(p, q)$) betrachtet werden, die von einer endlichen Anzahl von Parametern abhängen, damit der Bezug zu ökonomischen Anwendungen nicht verloren geht.

3.3.1 Schätzung der Kovarianzfunktion

Sei $X = \{X_t\}_{t \in \mathbb{Z}}$ ein linearer Prozess mit der Darstellung

$$X_t = \sum_{j=0}^{\infty} \gamma_j \delta_{t-j}, \quad t \in \mathbb{Z}, \quad \gamma_0 = 1, \quad |\gamma_j| \leq ce^{-aj}$$

für $a, c > 0$, $j \in \mathbb{N}$, $\mathbb{E} \delta_j = 0$, $\text{Var } \delta_j = \sigma^2$, $\{\delta_j\}$ unkorreliert .

Seien Beobachtungen X_0, \dots, X_n von X gegeben. Wie schätzen wir die Kovarianzfunktion $K(s)$ bzw. die Korrelationsfunktion $R(s)$?

Definieren wir folgenden Schätzer für $K(s)$:

$$\hat{K}_n(s) = \frac{1}{n} \sum_{j=0}^{n-s} X_j X_{j+s}, \quad s = 0, \dots, n . \quad (3.3.1)$$

Dementsprechend ist

$$\hat{R}_n(s) = \frac{\hat{K}_n(s)}{\hat{K}_n(0)}$$

ein Schätzer für $R(s)$.

Lemma 3.3.1

Unter den obigen Voraussetzungen an X und $\mathbb{E} \delta_t^4 < \infty$, $t \in \mathbb{Z}$ ist $\hat{K}(s)$ ein stark bzw. L^2 -konsistenter Schätzer für $K(s)$, d.h.:

1. $\hat{K}_n(s) \xrightarrow[n \rightarrow \infty]{\text{f.s.}} K(s)$, $s \in \mathbb{Z}$: $P(\lim_{n \rightarrow \infty} \hat{K}_n(s) = K(s)) = 1$ (starke Konsistenz)
2. $\hat{K}_n(s) \xrightarrow[n \rightarrow \infty]{L^2} K(s)$, $s \in \mathbb{Z}$: $\mathbb{E} |\hat{K}_n(s) - K(s)|^2 \xrightarrow[n \rightarrow \infty]{} 0$ (L^2 -Konsistenz)

Bemerkung 3.3.1

Wegen der Symmetrie von $K(s)$ und $R(s)$ genügt es, auch bei ihrer Schätzung positive s zu betrachten.

3.3.2 Parameterschätzung bei AR(p)-Prozessen

Sei $X = \{X_t\}_{t \in \mathbb{Z}}$ ein stationärer AR(p)-Prozess, der durch folgendes Gleichungssystem definiert ist:

$$\sum_{j=0}^p \beta_j X_{t-j} = \delta_t, \quad \forall t \in \mathbb{Z},$$

wobei $\beta_0 = 1$, $\beta_p \neq 0$, $\delta_t \sim \mathcal{N}(0, \sigma^2)$, $\{\delta_j\}$ unkorreliert, $\sigma^2 > 0$. Prozess X ist eindeutig durch β_1, \dots, β_p und σ^2 festgelegt. Wir stellen zwei Methoden zur Schätzung von β_1, \dots, β_p und σ^2 vor.

1. Lineare Regression

Aus der Definitionsgleichung von X folgt

$$X_t = - \sum_{j=1}^p \beta_j X_{t-j} + \delta_t, \quad t \in \mathbb{Z}. \tag{3.3.2}$$

Falls $X_0 = \dots = X_{1-p} = 0$ und die Anzahl n der beobachteten Größen X_1, \dots, X_n der Ungleichung $n > p$ genügt, dann kommt man durch mehrfache Anwendung von (3.3.2) für $t = 1, \dots, n$ zu folgendem linearen Gleichungssystem:

$$Y = A\beta + \delta,$$

wobei

$$Y = (X_1, \dots, X_n)^\top, \quad A = \begin{pmatrix} 0 & 0 & 0 & \dots & 0 \\ -X_1 & 0 & 0 & \dots & 0 \\ -X_1 & -X_2 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -X_1 & -X_2 & -X_3 & \dots & -X_p \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -X_{n-1} & -X_{n-2} & -X_{n-3} & \dots & -X_{n-p} \end{pmatrix},$$

$$\beta = (\beta_1, \dots, \beta_p)^\top, \quad \delta = (\delta_1, \dots, \delta_n)^\top, \quad \text{rg}(A) = p$$

Dies ist eine Regressionsgleichung mit Zielvariablen Y und Designmatrix A . Aus dem Appendix folgen MKQ-Schätzer für β und σ^2 :

$$\hat{\beta}_n = (A^\top A)^{-1} A^\top Y, \quad \hat{\sigma}_n^2 = \frac{1}{n-p} |Y - A\hat{\beta}|^2$$

Lemma 3.3.2

Falls (zusätzlich zu den obigen Annahmen) $\{\delta_t\}$ unabhängig sind und die Wurzeln des Polynoms

$$P(z) = \sum_{j=0}^p \beta_j z^j, \quad z \in \mathbb{C}$$

in $B_1(0) = \{z \in \mathbb{C} : |z| \leq 1\}$ liegen, dann sind $\hat{\beta}$ und $\hat{\sigma}^2$ stark konsistent:

$$\hat{\beta}_n \xrightarrow[n \rightarrow \infty]{\text{f.s.}} \beta, \quad \hat{\sigma}_n^2 \xrightarrow[n \rightarrow \infty]{\text{f.s.}} \sigma^2$$

2. Yule-Walker-Gleichungen

Hier wird vorausgesetzt, dass die Schätzung $\hat{R}(s)$ von der Autokorrelationsfunktion $R(s)$ aus Abschnitt 3.3.1 bereits vorliegt. Aus dem Satz 3.2.7 verwenden wir die Gleichungen (3.2.5)

$$R(p) = - \sum_{j=1}^p \beta_j R(p-j),$$

in denen R durch \hat{R} ersetzt wird. Es entsteht somit folgendes lineares Gleichungssystem für β :

$$\begin{pmatrix} \hat{R}(0) & \hat{R}(-1) & \dots & \hat{R}(1-p) \\ \vdots & \vdots & \ddots & \vdots \\ \hat{R}(p-1) & \hat{R}(p-2) & \dots & \hat{R}(0) \end{pmatrix} \cdot \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} = - \begin{pmatrix} \hat{R}(1) \\ \vdots \\ \hat{R}(p) \end{pmatrix}, \quad \hat{R}(-s) = \hat{R}(s).$$

Die Lösung des obigen Gleichungssystems ist ein Schätzer $(\hat{\beta}_1, \dots, \hat{\beta}_p)^\top$ von β . Die Schätzung von σ^2 erfolgt als $\hat{\sigma}^2 = \hat{K}(0)$. Auch hier gilt die Aussage des Lemmas 3.3.2 unter den selben Bedingungen.

Beispiel 3.2.1 (Umsätze im Einzelhandel - Fortsetzung)

Modellieren wir die stationäre Komponenten $X = \{X_t\}_{t \in \mathbb{Z}}$ der Zeitreihe $Z = \{Z_t\}_{t \in \mathbb{Z}}$, die die Umsätze im Einzelhandel darstellt, als AR(3)-Prozess; d.h. es sollen Parameter $\beta_1, \beta_2, \beta_3$ aus den Daten geschätzt werden. Mit Hilfe der Methoden 1 und 2 bekommt man

Methode	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
1)	-0.072	-0.079	5.77416
2)	-0.072	-0.078	5.77429

Tabelle 3.2.1d

4 Appendix

4.1 Regression

4.1.1 Einfache lineare Regression

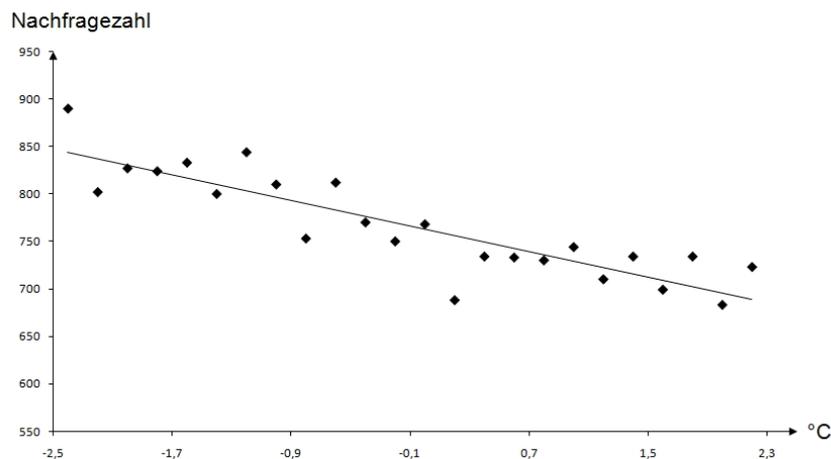
Betrachten wir das folgende lineare Modell

$$Y_i = \alpha + x_i\beta_i + \varepsilon_i, \quad i = 1, \dots, n,$$

wobei Y_i und ε_i , $i = 1, \dots, n$ Zufallsvariablen sind. x_1, \dots, x_n sind dagegen bekannte Konstanten.

Dabei kann z.B. Y_i die Nachfragezahl an Standheizungen in einem Jahr beschreiben, wobei x_i die jeweiligen Winterverhältnisse (durchschnittliche Wintertemperatur) angibt.

Man trägt die Wertepaare (x_i, Y_i) auf die Koordinatenebene auf:



Zielstellung: Schätze α und β aus den Daten Y_1, \dots, Y_n und x_1, \dots, x_n .

Bemerkung: Die Störgrößen $\varepsilon_1, \dots, \varepsilon_n$ sind nicht direkt beobachtbar.

Methode der kleinsten Quadrate



C. F. Gauß
(1777 - 1855)

Die Methode der kleinsten Quadrate geht auf den deutschen Mathematiker Carl Friedrich Gauß zurück, der dieses Verfahren bereits mit 18 Jahren entwickelte.

Berechne den *mittleren quadratischen Abstand* zu der Geraden $y = \alpha + x\beta$:

$$e(\alpha, \beta) = \frac{1}{n} \sum_{i=1}^n (Y_i - \alpha - \beta x_i)^2 \rightarrow \min_{\alpha, \beta}$$

Aus den notwendigen Bedingungen eines Extremums

$$e'_\alpha(\alpha, \beta) = e'_\beta(\alpha, \beta) = 0$$

folgen die Kleinsten-Quadrate-Schätzer $\hat{\alpha}$ und $\hat{\beta}$ für α und β :

$$\hat{\beta} = \frac{s_{xy}^2}{s_{xx}^2}, \quad \hat{\alpha} = \bar{Y}_n - \hat{\beta}\bar{x}_n, \quad (4.1.3)$$

wobei

$$s_{xx}^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_n)^2 \quad \text{die Strichprobenvarianz von } (x_1, \dots, x_n),$$

$$s_{yy}^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y}_n)^2 \quad \text{die Strichprobenvarianz von } (Y_1, \dots, Y_n),$$

$$s_{xy}^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_n)(y_i - \bar{y}_n) \quad \text{die Strichprobenkovarianz von } (x_1, \dots, x_n) \text{ und } (Y_1, \dots, Y_n)$$

und

$$\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i \quad \text{das arithmetische Mittel von } (x_1, \dots, x_n),$$

$$\bar{y}_n = \frac{1}{n} \sum_{i=1}^n y_i \quad \text{das arithmetische Mittel von } (Y_1, \dots, Y_n)$$

sind.

Mit ihrer Hilfe ist es nunmehr möglich, für ein vorgegebenes x (z.B. -20°C) den Wert Y , d.h. die Nachfragezahl an Standheizungen, zu prognostizieren, obwohl das oben genannte x nicht beobachtet wurde. Die prognostizierte Nachfrage ist gegeben durch:

$$Y = \hat{\alpha} + \hat{\beta}x.$$

Dabei ist es auch interessant zu testen, ob die Nachfrage nicht von den Winterverhältnissen abhängt:

$$\text{Haupthypothese } H_0 : \beta = 0 \quad \text{vs.} \quad \text{Alternative } H_1 : \beta \neq 0.$$

4.1.2 Multivariate lineare Regression

Die multivariate lineare Regression unterscheidet sich von der einfachen linearen Regression durch die Anzahl der Einflussgrößen x_i :

$$Y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_m x_{im} + \varepsilon_i, \quad i = 1, \dots, n, \quad m \ll n,$$

oder in Matrix-Form

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = \underbrace{\begin{pmatrix} x_{11} & \dots & x_{1m} \\ \vdots & & \vdots \\ x_{n1} & \dots & x_{nm} \end{pmatrix}}_X \underbrace{\begin{pmatrix} \beta_1 \\ \vdots \\ \beta_m \end{pmatrix}}_\beta + \underbrace{\begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}}_\varepsilon$$

bzw. $Y = X\beta + \varepsilon$ in Kurzschreibweise ,

wobei Y die *Zielvariablen* sind, X die $(n \times m)$ - *Designmatrix* mit $rg(X) \leq m \ll n$ und β der *Parameter-Vektor* ist.

Annahme:

- $\mathbb{E} \varepsilon_i = 0$ (kein systematischer Fehler bei der Messung von Y)
- $\varepsilon_i, \dots, \varepsilon_n$ unkorreliert: $\text{Cov}(\varepsilon_i, \varepsilon_j) = \delta_{ij} \sigma^2 = \begin{cases} \sigma^2 & , \quad i = j, \\ 0 & , \quad i \neq j. \end{cases}$
- $\text{Var} \varepsilon_i = \mathbb{E} \varepsilon_i^2 - (\mathbb{E} \varepsilon_i)^2 = \mathbb{E} \varepsilon_i^2 = \sigma^2, \quad i = 1, \dots, n$

Fragestellung:

Bestimme β aus den Daten Y und X . Dabei sind ε (und folglich Y) zufällig, X ist deterministisch. Die Fehler ε sind nicht direkt beobachtbar.

Wir bestimmen wie in Abschnitt 4.1.1 einen Schätzer $\hat{\beta}$ für β durch die Methode der kleinsten Quadrate:

Der mittlere quadratischer Fehler $e(\beta)$ soll dabei minimiert werden:

$$e(\beta) = \frac{1}{n} \sum_{i=1}^n (Y_i - \beta_1 x_{i1} - \dots - \beta_m x_{im})^2 \rightarrow \min_{\beta}$$

Ähnlich wie in Abschnitt 4.1.1 ist der MKQ-Schätzer $\hat{\beta}$ eine Lösung der sog. *Normalgleichung*

$$(X^T X)\beta = X^T Y .$$

Falls $rg(X) = m$, dann (und genau dann) ist $X^T X$ eine invertierbare Matrix und es gilt

$$\hat{\beta} = (X^T X)^{-1} X^T Y .$$

Falls $rg(X) < m$, dann ist $\hat{\beta}$ nicht eindeutig bestimmt. Dieser Fall wird hier nicht weiter betrachtet.

Der Schätzer $\hat{\beta}$ ist erwartungstreu für β , d.h. $\mathbb{E}\hat{\beta} = \beta \quad \forall \beta \in \mathbb{R}^m$.

Beweis

$\hat{\beta}$ ist ein linearer Schätzer bzgl. Y :

$$\begin{aligned} \mathbb{E}(\hat{\beta}) &= (X^T X)^{-1} X^T \mathbb{E}Y = (X^T X)^{-1} X^T \mathbb{E}(X\beta + \varepsilon) \\ &= \underbrace{(X^T X)^{-1} (X^T X)}_{\text{Id}} \beta + (X^T X)^{-1} X^T \underbrace{\mathbb{E}\varepsilon}_{=0} = \beta \end{aligned}$$

□

Weitere Fragestellungen:

Schätze die Varianz σ^2 der Werte Y :

$$\hat{\sigma}^2 = \frac{1}{n-m} |Y - X\hat{\beta}|^2 .$$

Der Schätzer $\hat{\sigma}^2$ ist ebenfalls erwartungstreu, d.h. $\mathbb{E}\hat{\sigma}^2 = \sigma^2$.

Weitere Ergebnisse zu den linearen Regressionsmodellen können den Skripten der Vorlesungen „Wirtschaftsstatistik“, „Statistik I“ und „Statistik II“ entnommen werden.

4.2 p-Wert

Definition 4.2.1

Der p-Wert ist das kleinste Signifikanzniveau, für das die Nullhypothese H_0 noch abgelehnt wird.

Beispiel 4.2.1

In neun verschiedenen amerikanischen Wintersportorten wurde während einer gewissen Beobachtungszeit die Anzahl der Besucher registriert. Es wird angenommen, dass diese

linear von der Gesamtlänge der zur Verfügung stehenden Pisten sowie der Liftkapazität abhängen.

Skigebiet	Pistenlänge	Liftkapazität	Besucherzahl
1	10.5	2200	19929
2	2.5	1000	5839
3	13.1	3250	23696
4	4.0	1475	9881
5	14.7	3800	30011
6	3.6	1200	7241
7	7.1	1900	11634
8	17.0	4200	36476
9	6.4	1850	12068

Tabelle 4.2.1

Es sei folgendes Regressionsmodell gegeben:

$$Y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} , \quad i = 1, \dots, 9$$

Teste die Hypothese H_0 , dass keine Abhängigkeit der Besucherzahl von den beiden Einflussgrößen (Pistenlänge, Liftkapazität) besteht für $\alpha = 0.05$:

$$H_0 : \beta_2 = \beta_3 = 0 \text{ vs. } H_1 : \beta_2 \neq 0 \text{ oder } \beta_3 \neq 0$$

Wir verwenden für den Hypothesentest die `summary`-Funktion des Statistik-Programms R. Der letzten Zeile des Outputs ist der p-Wert zu entnehmen.

Call:

```
lm(formula = Besucherzahl ~ 1 + Pistenlaenge + Liftkapazitaet)
```

Residuals:

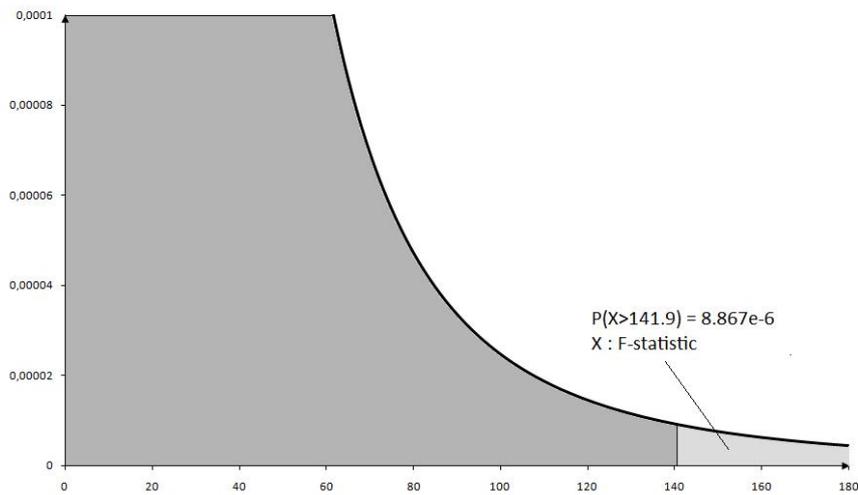
Min	1Q	Median	3Q	Max
-2418.2	-765.0	232.8	1110.7	2060.0

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-2020.700	1779.424	-1.136	0.299
Pistenlaenge	1098.981	678.282	1.620	0.156
Liftkapazitaet	4.227	3.087	1.369	0.220

Residual standard error: 1786 on 6 degrees of freedom

Multiple R-squared: 0.9793, Adjusted R-squared: 0.9724
F-statistic: 141.9 on 2 and 6 DF, p-value: 8.867e-06



Dichte der $F_{2,6}$ - Verteilung

Da der p-Wert = $8.867e-6$ ist, ist dies ein sehr stark signifikanter Hinweis darauf, dass H_0 verworfen werden sollte. Somit hat anscheinend mindestens einer der Faktoren (Pistenlänge, Liftkapazität) Einfluss auf die Besucherzahl.

Literaturverzeichnis

1. v. Auer, L. - Ökonometrie, Eine Einführung, Springer 2003
2. Greene, W.H. - Econometric Analysis, Prentice Hall, 2003
3. Fahrmeir, L., Kneib, T., Lang, S. - Regression: Modelle, Methoden und Anwendungen, Springer, 2007
4. Heil, J. - Einführung in die Ökonometrie, Oldenbourg, 1991
5. Judge, G.G. - The Theory and Practice of Econometrics, John Wiley & Sons, 1985
6. Kazmir, L. J. - Wirtschaftsstatistik, McGraw Hill, 1996
7. Löbus, J.-U. - Ökonometrie, Vieweg, 2001
8. Johnston, J. & DiNardo, J. - Econometric Methods, McGraw & Hill, 1997
9. Gujarati, D. N. - Basic Econometrics, McGraw & Hill, 2003
10. Hackl, P. - Einführung in die Ökonometrie, Pearson, 2005

Index

A	
AIC-Kriterium	46
asymptotische Tests	40
Autokorrelationsfunktion	63
Autokovarianzfunktion	63
B	
Bestimmtheitsmaß R^2	25
Bochner	
Satz von	67
Box-Cox-Test	28
Box-Cox-Transformation	18
Box-Cox-Verfahren	28
E	
exogene Variablen	5
Exponentialfamilie	31
F	
Fisher Scoring	39, 44
Fisher-Informationsmatrix	35, 45
G	
Gauß-Newton-Verfahren	11
H	
Hesse-Matrix	36
Hilbertraum	57
I	
Informationskoeffizient von Akaike	46
K	
Korrelationsfunktion	63
Kovarianzfunktion	63
L	
Likelihood-Ratio-Test	29, 30, 45
Lineare Approximation	8
Linkfunktion	31
natürliche	33
Log-Likelihood-Gleichungen	
konzentrierte	16
Logit-Modell	34, 43, 48
M	
Maximum-Likelihood-Methode	15, 20
Methode der kleinsten Quadrate	72
mittlerer quadratischer Abstand	72
Modelle	
quasilineare	7
verallgemeinerte lineare	31
N	
Newton-Verfahren	39
Normalengleichung	74
O	
Odd	34
P	
p-Wert	75
Paneldaten	4
Poisson-Modell	48
Poisson-Regression	38
Probit-Modell	34
Prozess	
autoregressiver, AR(1)	50
der gleitenden Mittel	59
linearer	56
Q	
Quantilfunktion der Normalverteilung	34
Querschnittsdaten	4
R	
Regression	
binäre kategoriale	34
einfache lineare	6, 72
logistische	34, 38
multivariate lineare	73
Regressoren	5
S	
saisonalen Anteil	53
Score-Funktion	45
Score-Statistik	45
Spektraldichte	67
Spektralfunktion	67
T	
Trend	53
W	
Wald-Statistik	45
Wold	
Zerlegungssatz von	60
Y	
Yule-Walker-Gleichung	66
Z	
Zeitreihen	4, 50, 53