ulm university universität
uulm

Ökonometrie

Vorlesungsskript

Prof. Dr. Evgeny Spodarev

Ulm Sommersemester 2011

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 und Herr Sebastian Fischer haben eine hervorragende Arbeit beim Tippen des Skriptes und bei der Erstellung zahlreicher Abbildungen, die den Text begleiten, geleistet. Dafür gilt ihnen mein herzlicher Dank.

Ulm, im November 2011 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 25

2.5.1 Reset-Tests 25

2.5.2 Welches Modell ist besser? - Bestimmtheitsmaß R^2 29

2.5.3 Box-Cox-Verfahren 31

2.5.4 Likelihood-Ratio-Test 33

2.6 Verallgemeinerte lineare Modelle 35

2.6.1 Maximum-Likelihood-Schätzung von β 40

2.6.2 Asymptotische Tests für β 45

2.6.3 Kriterien zur Modellwahl bzw. Modellanpassung 52

3 Zeitreihenanalyse **55**

3.1 Korrelierte Beobachtungen im Regressionsmodell 55

3.2 Zeitreihen 56

3.2.1 Autoregressive Prozesse: eine Einführung 57
3.2.2 Typen von Zeitreihen 60
3.2.3 Vorhersage von Zeitreihen 67
3.2.4 Eigenschaften der Zeitreihen 70
3.3 Statistik in der Zeitreihenanalyse 76
3.3.1 Schätzung der Kovarianzfunktion 76
3.3.2 Parameterschätzung bei AR(p)-Prozessen 77
3.3.3 Parameterschätzung bei ARMA(p,q)-Prozessen 79
3.3.4 Schätzung der Spektraldichte 81
3.3.5 Beispiele für die Schätzung der Spektraldichte 87
3.4 Zeitreihen mit stationärem Anteil 88
3.4.1 Zeitreihen mit polynomialem und saisonalem Anteil 92

4 Appendix **96**

4.1 Regression 96
4.1.1 Einfache lineare Regression 96
4.1.2 Multivariate lineare Regression 98
4.2 p-Wert 99

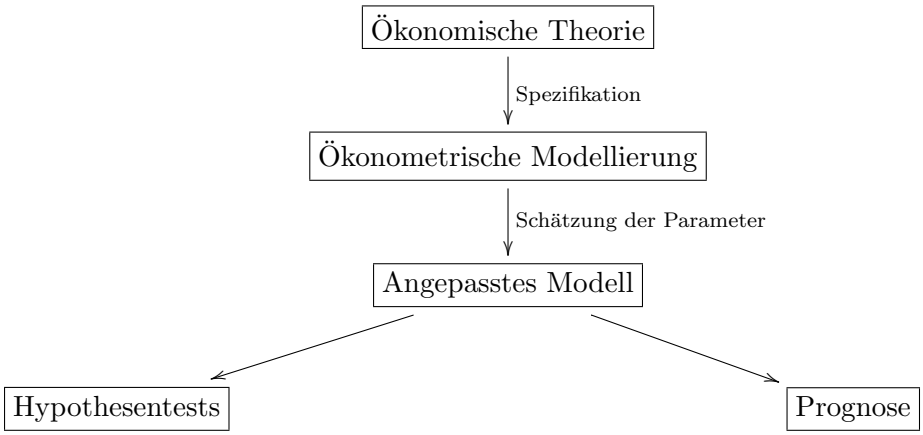
Literaturverzeichnis **102**

Index **103**

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 *ökonometrischen 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:

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

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 $x_{i1}, \ldots x_{im}, \ i = 1, \ldots, n$ 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

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

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.

$Y_i, i = 1, \ldots, 12$ - Milchleistung der Kuh i in l/Jahr
$x_i, i = 1, \ldots, 12$ - Kraftfutterabgabe in Zentner/Jahr

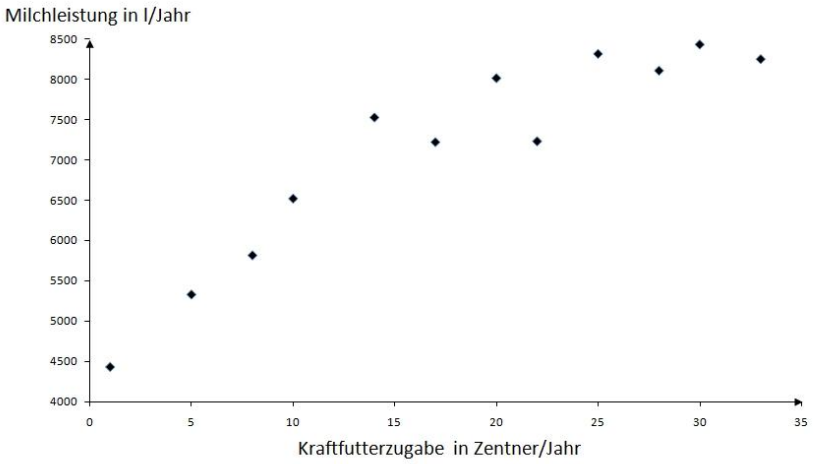
	1	2	3	4	5	6	7	8	9	10	11	12
Y_i	6525	8437	8019	8255	5335	7236	5821	7531	8320	4436	7225	8112
x_i	10	30	20	33	5	22	8	14	25	1	17	28

Tabelle 2.1.2

Einfache lineare Regression:

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

Plottet man die Punktewolke (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

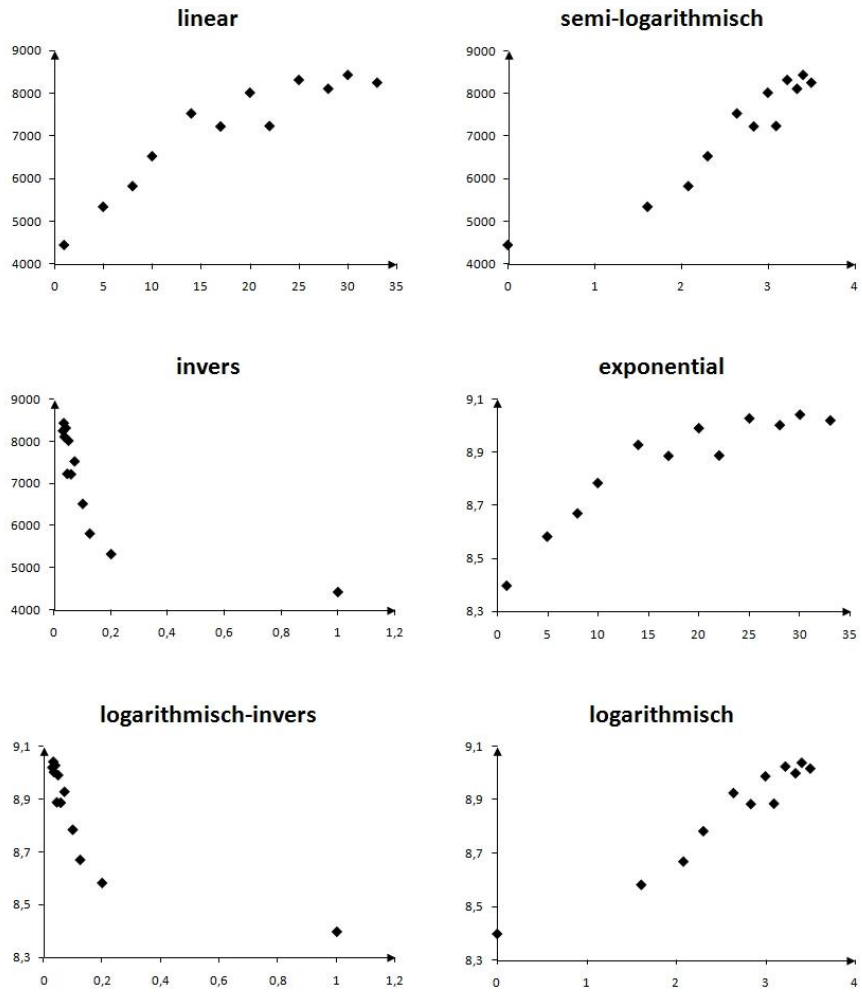
2.2.1 Quasilineare Modelle

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 Punktewolken 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 Punktewolken festzustellen.



Das logarithmische Modell passt am besten zu den Daten.

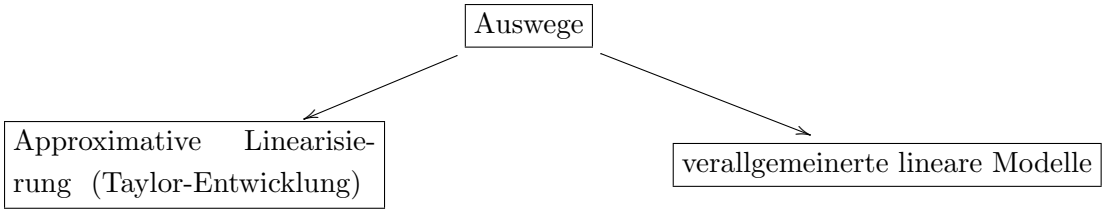
$$\underbrace{\log Y_i}_{Z_i} = \alpha + \beta \underbrace{\log x_i}_{t_i} + (\varepsilon_i) \quad \Leftrightarrow \quad Y_i = e^{\alpha} x_i^{\beta} e^{\varepsilon_i} \ , \quad i = 1, \ldots, 12$$

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:

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



2.2.2 Lineare Approximation

Ein essentiell nichtlineares (und kein quasilineares) Modell

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

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

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:

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

falls β in einer kleinen Umgebung von β^0 liegt. Der asymptotisch kleine Term $o(\left|\beta - \beta^0\right|)$ (für $\beta \to \beta^0$) 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:

$$\widetilde{x}_k = \frac{\partial f(\vec{x}, \beta)}{\partial \beta_k}, k = 1, \ldots, m \qquad \widetilde{x}_k^0 = \frac{\partial f(\vec{x}, \beta)}{\partial \beta_k}\Big|_{\beta=\beta^0}$$

$$\widetilde{x}_{i,k} = \frac{\partial f(\vec{x}_i, \beta)}{\partial \beta_k}, i = 1, \ldots, n \qquad \widetilde{x}_{i,k}^0 = \frac{\partial f(\vec{x}_i, \beta)}{\partial \beta_k}\Big|_{\beta=\beta^0}$$

$$\widetilde{x} = (\widetilde{x}_1, ..., \widetilde{x}_m)^\top \qquad \widetilde{x}^0 = (\widetilde{x}_1^0, ..., \widetilde{x}_m^0)^\top$$

$$\widetilde{X} = (\widetilde{x}_{i,k})_{\substack{i=1,\ldots,n \\ k=1,\ldots,m}} \qquad \widetilde{X}^0 = (\widetilde{x}_{i,k}^0)_{\substack{i=1,\ldots,n \\ k=1,\ldots,m}}$$

Dann gilt:

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

oder

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

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

Transformieren wir die unabhängigen Variablen Y_i nach der Regel

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

so bekommen wir ein neues lineares Modell

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

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 4.1.2 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

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

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:

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

Somit ist die lineare Approximation durch

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

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 $\sigma^2 = \mathrm{Var}\,(\varepsilon_i)$ des Modells zu schätzen? Nach wie vor setzen wir voraus, dass die Störgrößen unkorreliert sind mit

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

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

Der mittlere quadratische Fehler

$$
\begin{aligned}
e(\beta) &= \frac{1}{n}\sum_{i=1}^{n} (Y_i - f(\vec{x}_i,\beta))^2 = \frac{1}{n}\sum_{i=1}^{n} \hat{\varepsilon}_i^2 \ , \\
\hat{\varepsilon}_i &= Y_i - f(\vec{x}_i,\beta) \ , \quad i = 1,\dots,n
\end{aligned}
$$

wird gebildet und bezüglich β minimiert:

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

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

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

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

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

oder in gekürzter Matrix-Form

$$\widetilde{X}^\top \hat{\varepsilon} = 0$$

mit $\hat{\varepsilon} = (\hat{\varepsilon_1}, ..., \hat{\varepsilon_n})^\top$ (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 $\beta^1 = (\beta_1^1, ..., \beta_m^1)^\top$.

2. **Iterationsschritt**:
   Sei $\beta^k = (\beta_1^k, ..., \beta_m^k)^\top$ 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 $\beta^{k+1} = (\beta_1^{k+1}, \ldots, \beta_m^{k+1})^\top$ bezeichnet.

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

4. **Endergebnis**:
   Wähle das erste β^{k+1}, so dass $\left|\beta^{k+1} - \beta^k\right| \leq \delta$ 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

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

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

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:

	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

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

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

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

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

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

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

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

2.4 Nichtlineare Zusammenhänge in den Zielvariablen

Manchmal ist es vorteilhaft, Regressionszusammenhänge der Form

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

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 $\theta = (\theta_1, \ldots, \theta_k),\ k \geq 1$ 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:

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

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

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

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

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

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 , $\quad i = 1, \ldots, n$
- $g(y, \theta)$ ist invertierbar bzgl. y und $g^{-1}(y, \theta)$ stetig differenzierbar mit Ableitung $\neq 0$

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

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

wobei

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

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

Wiederholung des Dichte-Transformationssatzes *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

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

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

Die Likelihood-Funktion der Stichprobe $(Y_i, \ldots, Y_n)$ ist dann gegeben durch

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

und die Log-Likelihood-Funktion ist somit

$$\begin{aligned} &\log L(y_i, \ldots, y_n) = \\ &-\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\sigma^2 + \sum_{i=1}^{n} \log|J(y_i,\theta)| - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(g(y_i,\theta) - f(\vec{x}_i,\beta))^2 \ . \end{aligned} \tag{2.4.2}$$

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

$$\log L(y_i, \ldots, y_n) \to \max_{\theta,\beta,\sigma^2} \ . \quad .$$

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

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

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

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

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*:

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

Beispiel 2.4.1 (Verallgemeinerte Produktionsfunktion - Fortsetzung)

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

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) :

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

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

Vorgehensweise beim Optimierungsproblem:

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

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

   wobei M vorgegeben ist.

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

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

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

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

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

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

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? ⇒ *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

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

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

Gegeben sei ein lineares Regressionsmodell

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

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

$$Y_i^{(\theta)} = \sum_{j=1}^{m} x_{ij}^{(\lambda_j)}\beta_j + \varepsilon_i \ , \quad i = 1, \ldots, n \ , \tag{2.4.3}$$

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 $\theta \ , \ \lambda_1, \ldots, \lambda_m$ sind.

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

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

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

Spezialfälle:

$\theta = \lambda = 1 \Rightarrow$ lineares Modell

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

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

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

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

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 aus diesem Gitter G finde einen MKQ-Schätzer $\hat{\beta}(\lambda)$ für $\beta = (\beta_1, \ldots, \beta_m)$.

   Sei
   $$S(\lambda) = \sum_{i=1}^{n} \left(Y_i - \sum_{i=1}^{n} x_{ij}^{(\lambda)} \hat{\beta}_j(\lambda_o) \right)^2$$
   die Summe der quadrierten Residuen.
3. Wähle $\lambda_0 = \underset{\lambda \in G}{\operatorname{argmin}}\, S(\lambda)$. 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

Lineares Modell:

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

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

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

Log-lineares Modell:

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

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

Box-Cox-Transformiertes Modell:

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

λ - neuer Parameter

λ	0.30	0.40	0.41	0.42	...	0.46	0.47	0.48	0.49
$S(\lambda)$	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

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

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

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

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

weil

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

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

Die konzentrierte Log-Likelihood-Funktion ist dann gegeben durch

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

$\Rightarrow$ Um den ML-Schätzer $(\hat{\lambda}, \hat{\beta}) = \underset{\theta, \beta}{\operatorname{argmax}} \log L_c(y_1, \ldots, y_n | \lambda, \beta)$ 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 $\hat{\beta}(\lambda_0) = \underset{\beta}{\operatorname{argmax}} \log L_c(y_1, \ldots, y_n | \lambda_0, \beta)$
3. Gebe $\hat{\lambda} = \underset{\lambda_0 \in G}{\operatorname{argmax}} \log L_c(y_1, \ldots, y_n | \lambda_0, \hat{\beta}(\lambda))$ und $\hat{\beta} = \hat{\beta}(\hat{\lambda})$ als Ergebnis des Algorithmus aus

Beispiel 2.4.2 (Geldbedarf - Fortsetzung)

Wir betrachten das Modell

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

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 ökonometrischen Daten passt.

2.5.1 Reset-Tests

Wie kann man allgemein die Annahme der Linearität testen?

Beispiel der einfachen (linearen) Regression:

$$Y_i = \alpha + \beta x_i + \varepsilon_i \ , \quad i = 1, \ldots, n \tag{2.5.1}$$

$$Y_i = f(x_i) + \varepsilon_i \ , \quad i = 1, \ldots, n \tag{2.5.2}$$

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 $x = 0$:

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

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

$$x_i^2, x_i^3, x_i^4 : Y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_i^2 + \beta_3 x_i^3 + \beta_4 x_i^4 + \varepsilon_i, \quad i = 1, \ldots, n$$

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

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

auch alle Produkte der Form $x_1^{i_1}\cdot\ldots\cdot x_e^{i_e}$. 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 nicht-lineare Modell

$$Y_i = f(\vec{x}_i,\beta) + \varepsilon_i \ , \quad \text{wobei } \vec{x}_i = (x_{i1}, x_{i2})^\top \ , \quad i = 1,\ldots,n \ .$$

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

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

Für Testzwecke der Linearitätvon f bekämen wir somit (für $\vec{x}^0 = 0$) ein Modell der Form:

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

Deswegen wird folgende alternative Vorgehensweise angeboten:

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

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

Genauso hängt $\hat{Y}_i^k$ von x_i^k und ihren 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 $\hat{Y}_i^2,\ \hat{Y}_i^3,\ \hat{Y}_i^4, \ldots$:

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

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 $x_{i1}, \ldots, x_{im},\ i = 1, \ldots, n$, linear ist:

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

$$H_0 : f(\vec{x}, \beta) = \sum_{j=0}^{m} x_j \beta_j \text{ (Linearität) vs. } H_1 : f(\vec{x}, \beta) \neq \sum_{j=0}^{m} x_j \beta_j \text{ (keine Linearität) ,}$$

wobei $x_0 = 1$ angenommen wird.

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

$$Y_i = \sum_{j=0}^{m} x_{ij}\beta_j + \varepsilon_i \ , \quad i = 1, \ldots, n \ . \quad \text{Bilde } \hat{Y}_i = \sum_{j=0}^{m} x_{ij}\hat{\beta}_j \ , \quad i = 1, \ldots, n \ .$$

2. Ergänze das obige lineare Modell um 3 neue Ausgangsvariablen $\hat{Y}_i^2, \hat{Y}_i^3, \hat{Y}_i^4$:

$$Y_i = \sum_{j=1}^{m+3} x_{ij}\beta_j + \varepsilon_i, \quad i = 1, \ldots, n \ , \tag{2.5.3}$$

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 :

$$S = \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i\right)^2 = \sum_{i=1}^{n} \left(Y_i - \sum_{j=0}^{m} x_{ij}\hat{\beta}_j\right)^2$$

$$S^* = \sum_{i=1}^{n} \left(Y_i - \widetilde{Y}_i\right)^2 = \sum_{i=1}^{n} \left(Y_i - \sum_{j=0}^{m+3} x_{ij}\widetilde{\beta}_j\right)^2 \ ,$$

wobei $\widetilde{\beta}_j \ , \quad j = 1, \ldots, m+3$ die MKQ-Schätzer im neuen Modell (2.5.3) und $\widetilde{Y}_i = \sum_{j=1}^{m+3} x_{ij}\widetilde{\beta}_j \ , \quad j = 1, \ldots, n$ sind.
Es gilt $S^* \leq S$.

4. Bilde die *Teststatistik*

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

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: $T(Y_1, \ldots, Y_n, X) \sim F_{3,n-m-4}$

$\Big($F-Verteilung: $Z \sim F_{s,t}$ mit $s, t \in \mathbb{N} \Leftrightarrow Z \stackrel{\mathrm{d}}{=} \frac{\chi^2_s/s}{\chi^2_t/t}\Big)$

5. *Entscheidungsregel*: Lehne H_0 ab, falls

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

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)

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

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

$$Y_i = \beta_0 + \beta_1 x_i + \beta_2 \hat{Y}_i^2 + \beta_3 \hat{Y}_i^3 + \beta_4 \hat{Y}_i^4 + \varepsilon_i^*\ , \quad i = 1, \ldots, 12\ ,$$

wobei $\beta_0 = \alpha\ , \quad \beta_1 = \beta$ wird

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

getestet, wobei

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

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

Summe der quadrierten Residuen im ursprünglichen Modell:

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

Summe der quadrierten Residuen im ergänzten Modell:

$$S^* = \sum_{i=1}^{12} \left(Y_i - \widetilde{\beta}_0 - \widetilde{\beta}_1 x_i - \sum_{j=2}^{4} \widetilde{\beta}_j \hat{Y}_i^j \right)^2$$

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

$$Y_i = \sum_{j=1}^{m} x_i \beta_i + \varepsilon_i \ , \quad i = 1, \ldots, n$$

ein lineares Modell mit dem MKQ-Schätzer $\hat{\beta}_i$ für $\beta_j, \ j = 1, \ldots, m$.

Das *Bestimmtheitsmaß* R^2 ist gegeben durch

$$R^2 = \frac{\sum\limits_{i=1}^{n} \left(\hat{Y}_i - \overline{Y}_n \right)^2}{\sum\limits_{i=1}^{n} \left(Y_i - \overline{Y}_n \right)^2} \ ,$$

wobei

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

R^2 gibt den Anteil der Streuungsreduktion, die durch die Regression entsteht, aus der Gesamtstreuung der Stichprobe $(Y_1, \ldots, Y_n)$ 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 ?

$\Rightarrow$ 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:

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

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

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)?

G. E. P. Box
(1919 -)

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

1. Skalierung: Berechne das geometrische Mittel der Stichprobe $(Y_i, \ldots, Y_n)$

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

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

D. R. Cox
(1927 -)

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 (Box-Cox-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 $T \sim \chi^2_1$. Somit wird H_0 verworfen,
falls

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

wobei $\chi^2_{1,1-\alpha}$ das $(1-\alpha)$-Quantil der χ^2_1-Verteilung ist.

Beispiel 2.1.2 (Milchproduktion - Fortsetzung)

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

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

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:

$$T = \frac{12}{2} \Big| \log \frac{0.03807}{0.02873} \Big| \approx 1.68877 < \chi^2_{1,0.95} = 3.84146$$

$\Rightarrow$ 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)):

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

Sie enthält für

$\theta = \lambda = 1$	das lineare	} Modell als Spezialfall.
$\theta = 1, \lambda = 0$	das semi-logarithmische	
$\theta = 0, \lambda = 1$	das Exponential-	
$\theta = 0, \lambda = -1$	das log-inverse	
$\theta = 1, \lambda = -1$	das inverse	

Natürlich sind auch andere Werte für θ und λ denkbar.

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

1. Skaliere die Variablen Y_i durch das geometrische Mittel $\widetilde{Y}_n$ von $Y_1, \ldots, Y_n$:

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

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

$$Y_i^{*(\theta)} = \sum_{j=1}^{m} \beta_j x_{ij}^{(\lambda)} + \varepsilon_i \ , \quad i = 1, \ldots, n \tag{2.5.4}$$

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 (2.5.4). Auf Basis dieser Schätzung errechne die Summe der Residuenquadrate

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

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

2.5.4 Likelihood-Ratio-Test

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

Es ist folgendes allgemeines ökonometrisches Modell gegeben:

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

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:

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

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

Es gilt

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

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

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

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

Die Hypothese H_0 wird abgelehnt, falls

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

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. $\alpha = 0.01, 0.05, \ldots$).

Bemerkung 2.5.1

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

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

ist nicht χ_1^2-verteilt - sie ist nicht mehr nichtnegativ, weil die MKQ-Schätzer $\hat{\beta}$ und $\widetilde{\beta}$ in beiden Modellen (i.A.) nicht übereinstimmen.

Beispiel 2.4.2 (Geldbedarf - Fortsetzung)

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

Problem:

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

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

Nach der Berechnung gilt:

$$\begin{aligned}
&\log L(y_1, \ldots, y_n, X|\hat{\lambda}, \hat{\beta}) = -116.51 \quad (\hat{\lambda} = -0.35) \\
&\log L(y_1, \ldots, y_n, X|0, \hat{\beta}) = -118.073 \quad (\lambda = 0) \\
&\log L(y_1, \ldots, y_n, X|1, \hat{\beta}) = -130.1333 \quad (\lambda = 1)
\end{aligned}$$

$$\begin{aligned}
T_{\lambda=0}(y_1, \ldots, y_n, X) &= -2(-118.073 + 116.51) = 3.13 \\
T_{\lambda=1}(y_1, \ldots, y_n, X) &= -2(-130.133 + 116.51) = 27.25
\end{aligned}$$

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

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

Fazit:

Für die ökonometrische 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 ökonometrischen 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 $\beta = (\beta_1, \ldots, \beta_m)^\top$ 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 $Y_1, \ldots, Y_n$ eine Zufallsstichprobe der Zielvariablen des Modells und sei $X = (x_{ij})_{\substack{i=1,\ldots,n \\ j=1,\ldots,m}}$ die Designmatrix der Ausgangsvariablen, die hier nicht zufällig sind.

Definition 2.6.1

Das *verallgemeinerte lineare Modell* ist gegeben durch

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

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

Unter der Annahme, dass g explizit bekannt ist, soll hier der Parametervektor β aus $(Y_1,\ldots,Y_n)$ geschätzt werden. Wir setzen voraus, dass $Y_i\ , i=1,\ldots,n$, 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 $a:\mathbb{R}\times\mathbb{R}_+\to\mathbb{R}$ und $b:\Theta\to\mathbb{R}$ gibt, für die

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

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

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

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

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

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

  bzw.

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

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

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

   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

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

   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 $b : \Theta \to \mathbb{R}$ 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 $Y_i,\ i = 1, \ldots, n$ seien also unabhängig verteilt mit einer Verteilung, die zur Exponetialfamilie gehört und einer (Zähl)Dichte wie in (2.6.2) bzw. (2.6.3). Setzen wir voraus, dass $b : \Theta \to \mathbb{R}$ 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 $g : G \to \mathbb{R}$ 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 $(\theta_1, \ldots, \theta_n)^\top = X\beta$.

Beweis

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

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

Da g invertierbar ist, gilt

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

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

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

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

□

Beispiel 2.6.2

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

1. **Normalverteilung**: da $b(\mu) = \frac{\mu^2}{2}$, 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

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

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

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

Das Verhältnis

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

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

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

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 Linkfunktion. Mit ihrer Hilfe bekommt man das sog. *Probit-Modell*:

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

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

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

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 $Y = (Y_1, \ldots, Y_n)$ in folgender Form aufschreiben:

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

Aus dem Beweis des Lemmas 2.6.2 folgt, dass

$$\theta_i = (b')^{-1}(g^{-1}(x_i^\top \beta)) \ , \quad i = 1, \ldots, n \ , \tag{2.6.5}$$

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:

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

Dafür wird die notwendige Bedingung des Extremums

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

untersucht. Verwenden wir folgende Bezeichnungen:

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

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

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

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

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

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

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

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

Ein Beweis befindet sich im *„Stochastik III“ - 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:

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

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

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

und somit

$$U_j(\beta) = \frac{1}{\tau^2} \sum_{i=1}^n x_{ij} \left(Y_i - \mu_i(\beta)\right) \ , \quad j = 1, \ldots, m \ ,$$
$$I_{jk}(\beta) = \frac{1}{\tau^4} \sum_{i=1}^n x_{ij} x_{ik} \sigma_i^2(\beta) \ , \quad j, k = 1, \ldots, m \ .$$

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

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

von Bedeutung sein werden.

Satz 2.6.2

Es gilt

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

wobei

$$u_i = \frac{\partial g^{-1}(\eta_i)}{\partial \eta_i} \quad \text{und} \quad \nu_i = \frac{1}{\tau^2} \cdot \frac{\partial^2((b')^{-1} \circ g^{-1}(\eta_i))}{\partial \eta_i^2} ,$$
$$\mu_i(\beta) = \mathbb{E}\, Y_i , \quad \sigma_i^2(\beta) = \mathrm{Var}\,(Y_i) , \quad \eta_i = x_i^\top \beta , \quad i = 1, \ldots, n .$$

Beweis

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

$$\begin{aligned}
W_{jk}(\beta) = \frac{\partial}{\partial\beta_k} U_j(\beta) &\overset{\text{Satz 2.6.1}}{=} \frac{\partial}{\partial\beta_k} \sum_{i=1}^{n} x_{ij} \left(Y_i - \mu_i(\beta)\right) \frac{\partial g^{-1}(\eta_i)}{\partial \eta_i} \frac{1}{\sigma_i^2(\beta)} \\
&= \sum_{i=1}^{n} x_{ij} \left((Y_i - \mu_i(\beta)) \frac{\partial}{\partial\beta_k} \left(\frac{\partial g^{-1}(\eta_i)}{\partial \eta_i} \frac{1}{\sigma_i^2(\beta)} \right) - \frac{\partial g^{-1}(\eta_i)}{\partial \eta_i} \frac{1}{\sigma_i^2(\beta)} \frac{\partial \mu_i(\beta)}{\partial \beta_k} \right) \\
&= \sum_{i=1}^{n} \Bigg(x_{ij} (Y_i - \mu_i(\beta)) \frac{\partial}{\partial\beta_k} \left(\frac{b''((b')^{-1}(g^{-1}(\eta_i)))((b')^{-1} \circ g^{-1})'(\eta_i)}{\tau^2 b''((b')^{-1}(g^{-1}(\eta_i)))} \right) \\
&\quad - \left(\frac{\partial g^{-1}(\eta_i)}{\partial \eta_i} \right)^2 \frac{x_{ik}}{\sigma_i^2(\beta)} \Bigg) \\
&= \sum_{i=1}^{n} x_{ij} x_{ik} \left((Y_i - \mu_i(\beta))\nu_i - u_i^2 \frac{1}{\sigma_i^2(\beta)} \right) .
\end{aligned}$$

□

Für verallgemeinerte lineare Modelle mit natürlichen Linkfunktionen gilt insbesondere

$$W(\beta) = -I(\beta) = -\frac{1}{\tau^4}\sum_{i=1}^{n} x_{ij}x_{ik}\sigma_i^2(\beta) \ , \tag{2.6.6}$$

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

   $$\begin{aligned} U(\beta) &= \frac{1}{\sigma^2}X^\top(Y - X\beta) \ , \\ I(\beta) &= \left(\mathbb{E}\left(U_i(\beta)U_j(\beta)\right)\right)_{i,j=1,\ldots,m} = \frac{1}{\sigma^2}X^\top X \ , \\ W(\beta) &= -I(\beta) \ . \end{aligned}$$

2. **Logistische Regression**: hier gilt $\tau^2 = 1,\ \mu_i = p_i,\ \sigma_i^2 = p_i(1-p_i)$, $i = 1, \ldots, n,\ p_i \in (0,1)$ und somit

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

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

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

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

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

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?

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

Eine hinreichende Bedingung eines Maximums ist, dass die Hesse-Matrix $W(\beta)$ negativ definit ist.

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 $W(\beta) = -\frac{1}{\tau^4} X^\top diag(\sigma_i^2(\beta)) X$ ist negativ definit, falls $rg(X) = m$ und $0 < \sigma_i^2(\beta) < \infty$ für alle $i = 1, \ldots, n$.
  Unter diesen Bedingungen existiert also ein eindeutiger 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 $\hat{\beta}_k$, $k = 0, 1, \ldots$ auf folgende Art und Weise:
   - Nehme die Taylor-Entwicklung von $U(\beta)$ bis zur ersten Ordnung an der Stelle
     $$\hat{\beta}_k : U(\beta) \approx U(\hat{\beta}_k) + W(\hat{\beta}_k)(\beta - \hat{\beta}_k) \tag{2.6.7}$$
   - Setze sie gleich Null: $U(\hat{\beta}_k) + W(\hat{\beta}_k)(\beta - \hat{\beta}_k) = 0$
   - Die Lösung dieses Gleichungssystems ist $\hat{\beta}_{k+1}$:
     $$\hat{\beta}_{k+1} = \hat{\beta}_k - W^{-1}(\hat{\beta}_k) \cdot U(\hat{\beta}_k) \ , \quad k = 0, 1, 2, \ldots \ ,$$
     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

$$\mathbb{E}\, W(\beta) = -I(\beta) \tag{2.6.8}$$

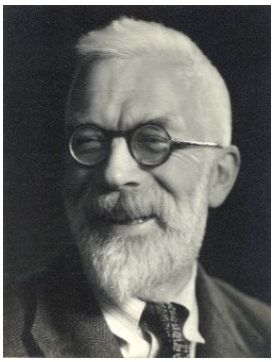
ersetzt wird. Dass die Identität (2.6.8) stimmt, folgt aus Satz 2.6.2, Formel (2.6.6) und der Tatsache, dass $\mathbb{E}\, Y_i = \mu_i,\ i = 1, \ldots, n$. Wenn man voraussetzt, dass $rg(X) = m$ und $u_i \neq 0,\ i = 1, \ldots, n$, so ist nach Satz 2.6.1 $I(\beta)$ invertierbar. Dieses Verfahren wird *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

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

einsetzt.
Im Falle einer natürlichen Linkfunktion gilt nach Bemerkung 2.6.1



R.A. Fisher
(1890 - 1962)

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

[1]

2.6.2 Asymptotische Tests für β

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

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

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 $Y = (Y_1, \ldots, Y_n)^\top$ von einigen Ausgangsvariablen (z.B. $(x_{1j}, \ldots, x_{nj})^\top$ im Falle der Hypothese $\beta_j = 0$) unabhängig sind.

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

Definition 2.6.5

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

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

1. $Z_n \xrightarrow[n\to\infty]{d} Z$ (Konvergenz in Verteilung gegen einen Zufallsvektor Z), falls $P(Z_n \le x) \xrightarrow[n\to\infty]{d} P(Z \le x) \quad \forall\, x = (x_1, \dots, x_n)^\top \in \mathbb{R}^n$, wobei $\{Z_n \le x\}$ bedeutet, dass $\{Z_{nj} \le x_j \ , \quad j = 1, \dots, k\}$ für $Z_n = (Z_{n1}, \dots, Z_{nk})^\top$.
2. $Z_n \xrightarrow[n\to\infty]{P} Z$ (Konvergenz in Wahrscheinlichkeit gegen einen Zufallsvektor Z), falls $P(|Z_n - Z| > \varepsilon) \xrightarrow[n\to\infty]{} 0 \quad \forall\, \varepsilon > 0$, 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 schwach konsistente Folge von ML-Schätzern für β: $U(\hat\beta_n) = 0, \ n \in \mathbb{N}$ und $\hat\beta_n \xrightarrow[n\to\infty]{P} \beta$.
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 $\lim\limits_{n\to\infty} \Gamma_n = 0$ und $\lim\limits_{n\to\infty} \Gamma_n^\top I_n(\beta)\Gamma_n = K^{-1}(\beta)$, wobei $K(\beta)$ eine symmetrische positiv definite $(m \times m)$-Matrix ist.

Satz 2.6.3

Unter obigen Voraussetzungen gilt:

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

Bemerkung 2.6.2

1. Falls $I_n(\beta)$ positiv definit ist und $\lim\limits_{n\to\infty} I_n^{-\frac{1}{2}}(\beta) = 0$ (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 $\mathrm{Id} = \mathrm{diag}\,(1, \dots, 1)$:

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

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

In diesem Fall lautet die Aussage 1 des Satzes 2.6.3

$$(I_n^{\frac{1}{2}})^\top(\beta)(\hat{\beta}_n - \beta) \xrightarrow[n\to\infty]{d} \mathcal{N}(0, \mathrm{Id}) \ . \tag{2.6.9}$$

2. Betrachten wir den Spezialfall des logistischen Regressionsmodells:
   $Y_i \sim$ Bernoulli $(p_i),\ i = 1, \dots, n.$

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

   Aus Beispiel 2.6.3(2) ist

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

   Falls $p_i \in (0, 1)$ und $\inf\limits_i p_i(1 - p_i) > 0,\ rg(X) = m$, dann ist $I_n(\beta)$ positiv definit. Falls zusätzlich $\lim\limits_{n\to\infty} (X^\top X)^{-1} = 0$, dann gilt $\lim\limits_{n\to\infty} I_n^{-\frac{1}{2}}(\beta) = 0$ und somit befinden wir uns in der Situation des Punktes 1 dieser Bemerkung.

Wie verwendet man nun den Satz 2.6.3 zum Testen der Hypothesen

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

oder komponentenweise

$$H_0 : \beta_j = \beta_{j0} \ ,\ j = 1, \dots, m \ \text{ vs. } \ H_1 : \exists j_1 : \beta_{j_1} \neq \beta_{j_1 0} \quad ?$$

Sei

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

ein verallgemeinertes lineares Modell mit natürlicher Linkfunktion g.

Nach Bemerkung 2.6.1 gilt

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

Deshalb gilt

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

Bei Vorgabe eines Exponential-Modells (τ, b - bekannt), der Stichprobe der Zielvariablen Y und der Designmatrix X wird H_0 verworfen, falls $T_n > \chi^2_{m,1-\alpha}$, wobei m die Anzahl der Parameter im Modell, $\chi^2_{m,1-\alpha}$ das $(1 - \alpha)$-Quantil der χ^2_m - 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 $n \to \infty$) 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 aus der Teststatistik T_n^* abgeleitete Teststatistik T_n^1. H_0 wird verworfen, falls

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

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

$$P_{H_0}(|T_n^1| > z_{1-\frac{\alpha}{2}}) = 1 - P_{H_0}(|T_n^*| \leq z_{1-\frac{\alpha}{2}}) \xrightarrow[n\to\infty]{} 1 - \Phi(z_{1-\frac{\alpha}{2}}) + \underbrace{\Phi(-z_{1-\frac{\alpha}{2}})}_{1-\Phi(z_{1-\frac{\alpha}{2}})}$$

$$= 1 - \left(1 - \frac{\alpha}{2}\right) + 1 - \left(1 - \frac{\alpha}{2}\right) = \alpha \ ,$$

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:

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

x_{i2} - 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} - Zahlungsverhalten beim Kunden : $= \begin{cases} 1 \ , & \text{gut} \\ 0 \ , & \text{sonst} \end{cases}$

x_{i6} - 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	$Y=1$	$Y=0$
	$0 < \ldots \leq 500$	1.00	2.14
	$500 < \ldots \leq 1000$	11.33	9.14
	$1000 < \ldots \leq 1500$	17.00	19.86
	$1500 < \ldots \leq 2500$	19.67	24.57
	$2500 < \ldots \leq 5000$	25.00	28.57
	$5000 < \ldots \leq 7500$	11.33	9.71
	$7500 < \ldots \leq 10000$	6.67	3.71
	$10000 < \ldots \leq 15000$	7.00	2.00
	$15000 < \ldots \leq 20000$	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

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

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

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

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

$$\log \frac{p_i}{1-p_i} = \beta_0 + x_{i1}\beta_1 + x_{i2}\beta_2 + x_{i3}\beta_3 + x_{i4}\beta_4 + x_{i5}\beta_5 + x_{i6}\beta_6 \quad \text{für } i = 1, \ldots, n\ ,$$
$$\text{wobei } \beta = (\beta_0, \ldots, \beta_6)^\top\ , \quad m = 7\ .$$

Ziel: Schätze $\beta_0, \ldots, \beta_6$ und prüfe, welche Faktoren für die künftige Kreditvergabe relevant sind.

Hier wird $\sqrt{(I_n^{-1}(\hat{\beta}))_{ii}}$ als asymptotische Standardabweichung von $\hat{\beta}_i$ interpretiert (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 Intui-

	Wert	$\sqrt{(I_n^{-1}(\hat{\beta}))_{ii}}$	T_n^1	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

tion 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^1	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

$H_0 : \beta_3^2 = 0$ (lineares Modell) vs. $H_1 : \beta_3^2 \neq 0$ (quadratisches Modell) bzw.
$H_0 : \beta_4^2 = 0$ (lineares Modell) vs. $H_1 : \beta_4^2 \neq 0$ (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 $C = \mathrm{Id}$, $d = \beta_0$. Die neuen Hypothesen beinhalten Aussagen über die Linearkombinationen der Parameterwerte. Wie soll H_0 vs. H_1 getestet werden?

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

Sei $\hat{\beta}_n$ der ML-Schätzer von β unrestringiert, d.h. $\hat{\beta}_n = \underset{\beta \in \mathbb{R}^m}{\operatorname{argmax}} \log L(Y, \beta)$.
Die Idee der folgenden Tests ist es, $\widetilde{\beta}_n$ mit $\hat{\beta}_n$ zu vergleichen. Falls die Abweichung $\hat{\beta}_n - \widetilde{\beta}_n$ groß ist, soll H_0 abgelehnt werden.

Satz 2.6.4

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

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

Wir führen folgende Teststatistiken ein:

1. **Likelihood-Ratio-Teststatistik**:
$$\widetilde{T}_n = 2(\log L(Y, \hat{\beta}_n) - \log L(Y, \widetilde{\beta}_n))$$
2. **Wald-Statistik**:
$$\widetilde{T}_n^* = (C\hat{\beta} - d)^\top (C I_n^{-1}(\widetilde{\beta}_n) C^\top)^{-1} (C\hat{\beta} - d)$$
3. **Score-Statistik**:
$$\overline{T}_n^* = U(\widetilde{\beta}_n)^\top I_n^{-1}(\widetilde{\beta}_n) U(\widetilde{\beta}_n)$$

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

$$\widetilde{T}_n \xrightarrow[n \to \infty]{d} \chi^2_m \ .$$

Folgerung 2.6.1

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

$$\widetilde{T}_n(\widetilde{T}_n^*, \overline{T}_n) > \chi^2_{m,1-\alpha} \ .$$

Dies ist ein asymptotischer Test zum Signifikanzniveau α.

Beispiel 2.6.4 (Fortsetzung)

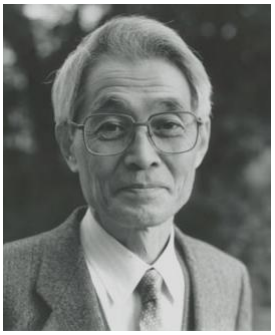
Es ergeben sich folgende Werte für die Teststatistiken:

$\widetilde{T}_n = 12.44$, p-Wert: 0.0020
$\widetilde{T}_n^* = 11.47$, p-Wert: 0.0032 .

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:

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

wobei $Y = (Y_1, \ldots, Y_n)$ 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:

$$\text{Lineares Modell} : \mathrm{AIC} = 1043.815$$
$$\text{Quadratisches Modell} : \mathrm{AIC} = 1035.371$$

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

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

ein verallgemeinertes lineares Modell mit Linkfunktion g und Parametervektor $\beta = (\beta_1, \ldots, \beta_m)^\top$. Teilen wir die Zielvariablen $Y_1, \ldots, Y_n$ 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:

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

In die Gruppe l fallen alle Beobachtungen Y_i, die zu $(a_l, b_l]$ gehören. Dabei müssen $(a_l, b_l]$ so gewält 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,
- $\overline{Y}_l = \frac{1}{n_l} \sum_j Y_j$ 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:

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

$$D_n = -2\tau^2 \sum_{l=1}^{k} \left(l_l(\hat{\mu}_l) - l_l(\overline{Y}_l) \right)$$

Satz 2.6.5

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

$$K_n \xrightarrow[n\to\infty]{d} \chi^2_{k-m-1}$$

$$D_n \xrightarrow[n\to\infty]{d} \chi^2_{k-m-1}$$

[3] Dies 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.

Folgerung 2.6.2

Mit Hilfe der Behauptungen des Satzes 2.6.5 können die Hypothesen

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

vs.

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

folgendermaßen getestet werden:

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

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

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**: $Y_i \sim \text{Bernoulli}(p_i),\ i = 1, \ldots, n$

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

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

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

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

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

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

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

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 ($t = 1, \ldots, 24$) zu analysieren. Es wird versucht, ein Modell der einfachen linearen Regression

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

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 Modellanpasung 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

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

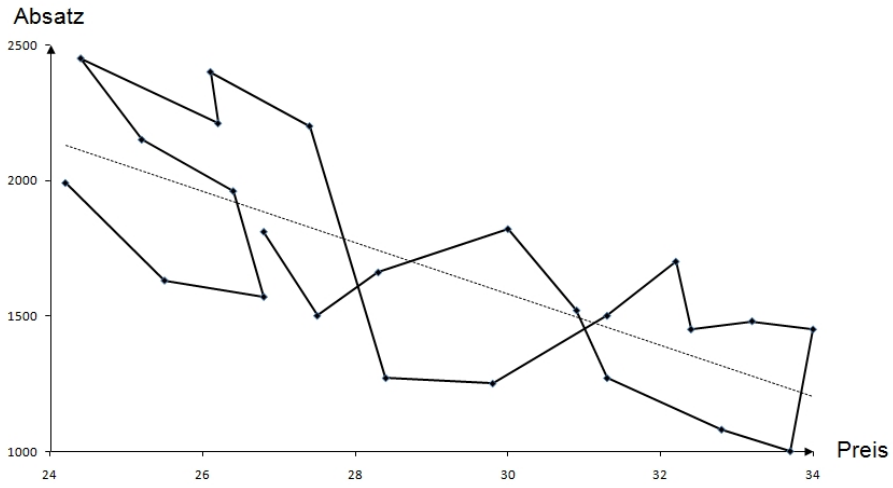
leicht zu berechnen sind, sind sie nutzlos, falls weitere Güteaussagen fehlen.

Bei der grafischen Darstellung der Punktewolke $(x_t, Y_t),\ t = 1, \ldots, 24,$ 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 $t+1,\ t+2, \ldots$.

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:

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

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

$$\text{Somit liegt bei } \rho \left\{ \begin{array}{ll} > 0 & \text{positive} \\ = 0 & \text{keine} \\ < 0 & \text{negative} \end{array} \right| \text{ Korrelation zwischen } Z_t \text{ und } Z_{t-k} \text{ vor.}$$

Beweis

1. Es gilt

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

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.

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

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

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

Somit gilt

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

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

$$\varepsilon_0 = 0\ , \quad \varepsilon_t = \rho\varepsilon_{t-1} + \delta_t\ , \quad t \in \mathbb{N}\ , \quad \delta_t \sim \mathcal{N}(0, \sigma_\delta^2)\ \text{u.i.v.}\ , \quad |\rho| < 1\ .$$

Wie kann das Modell

$$Y_t = \alpha + \beta x_t + \varepsilon_t\ , \quad t = 1, \ldots, n\ , \tag{3.2.1}$$

als einfache lineare Regression mit unabhängigen Störgrößen dargestellt werden?

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

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

Nach der Substitution

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

bekommen wir wieder ein einfaches lineares Regressionsmodell

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

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 $\hat{\alpha}^{(i)}, \ \hat{\beta}^{(i)}, \ i = 1, 2$:

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

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:

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

Praktisch aber kann das Modell (3.2.2) nicht direkt eingesetzt werden, weil ε_t und somit $\rho, \ \delta_t$ 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:

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

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

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

- 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 3.2.1

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

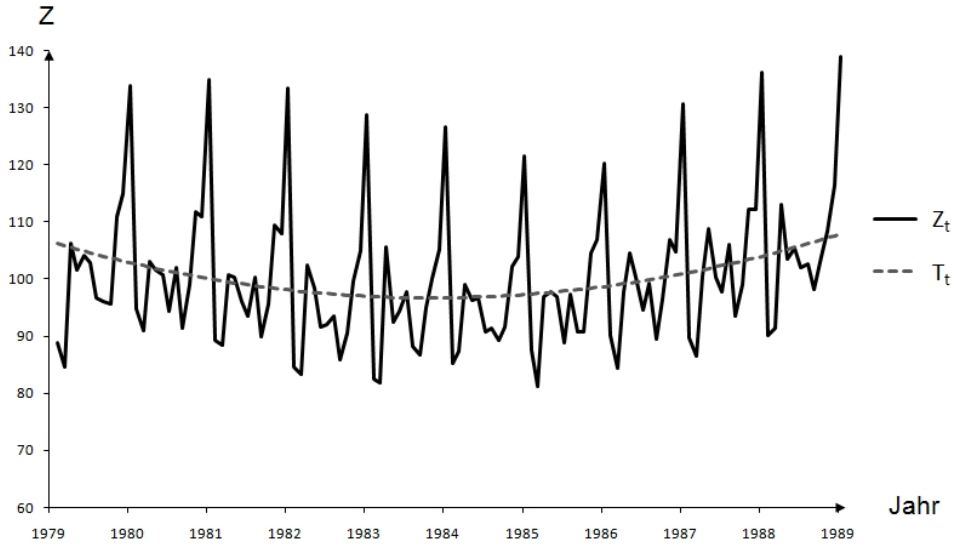
Jahr	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988
Preisindex	78.3	82.8	88.1	92.8	95.9	98.5	100.5	100.0	100.6	102.0

Tabelle 3.2.2

Die zu analysierende Zeitreihe ist nun

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

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



Zeitreihen Z_t und ihr Trend T_t

$$\text{Trend: } T_t = 0.00295t^2 - 0.342t + 106.555 \ , \quad t \in \mathbb{Z} \ ,$$
$$\text{Saisonaler Anteil: } S_t = S_t^{(1)} + S_t^{(2)} \ , \quad t \in \mathbb{Z} \ ,$$

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

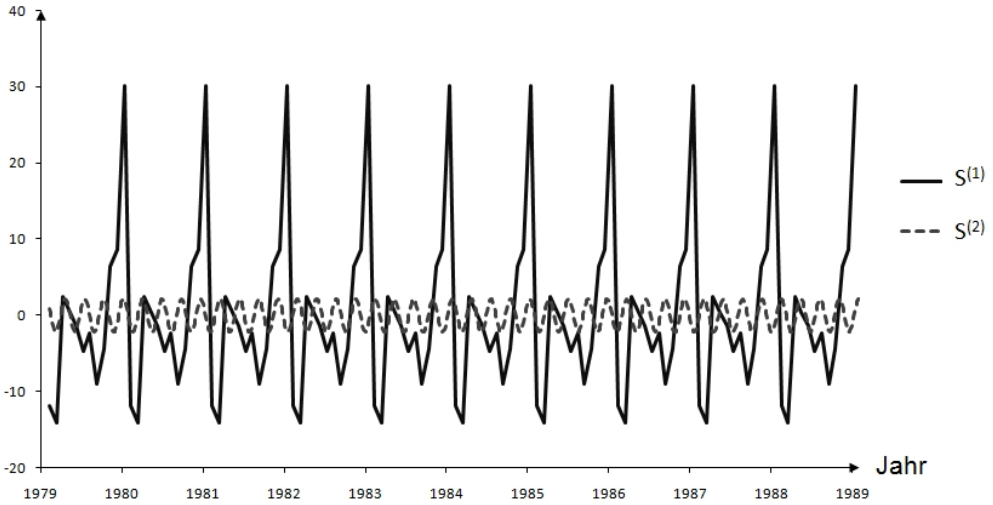
Tabelle 3.2.3 - Werte von $S_t^{(1)}$

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

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

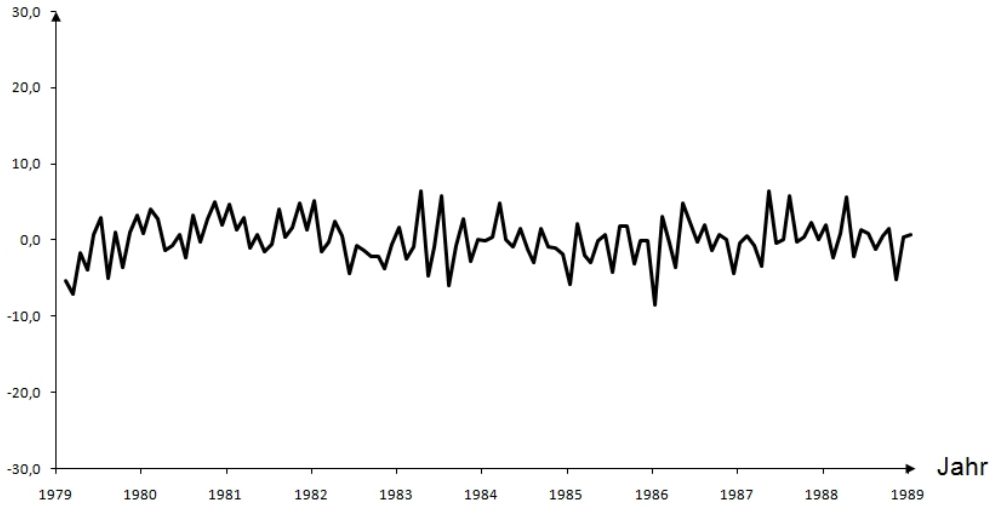
gegeben. Sie besitzt die Periode $p^{(2)} = \frac{2\pi}{2.2} = 2.86$.

Der stationäre Anteil der Zeitreihe Z ergibt sich als Differenz



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

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



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) $\mathbb{E}\, X_t^2 < \infty \quad \forall\, t \in \mathbb{Z} \ , \quad \mathbb{E}\, X_t = const.$

   b) $\mathrm{Cov}\,(X_t, X_{s+t}) = \mathrm{Cov}\,(X_0, X_s)$ hängt nicht von $t \in \mathbb{Z}$ ab.

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

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

d.h. $P(X_{t+t_1} \in B_1, \ldots, X_{t+t_n} \in B_n)$ 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

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

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

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 $\sum\limits_{i=0}^{\infty} \gamma_i \delta_{t-i}$ zu verstehen?

Führen wir die Zufallsvariablen

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

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 $(\Omega, \mathcal{F}, P)$ ist, d.h. ein vollständiger linearer normierter unendlich dimensionaler Vektorraum.

- $L^2 = \{Y -$ Zufallsvariable auf $(\Omega, \mathcal{F}, P) : \mathbb{E}\,Y^2 < \infty$ 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 $L^2(\Omega, \mathcal{F}, P)$

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

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

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

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.

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

Beweis

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

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

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

   für $\{a_n\} : a_n \to 0$ folgt $(a_n, b) \to 0,$ wobei$\{a_n\}$ eine Folge in einem Hilbertraum H mit Skalarprodukt $(\cdot,\cdot)$ ist . Deshalb gilt $\mathbb{E}\, X_t \equiv 0\ ,\quad \forall\ t \in \mathbb{Z}$.

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

   O.B.d.A. setzen wir $s \in \mathbb{N}_0$. Es gilt

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

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

$$
\begin{aligned}
|\mathbb{E}\,(X_t X_{t+s}) - \mathbb{E}\,(X_t^n X_{t+s}^n)| &= |\mathbb{E}\,(X_t X_{t+s}) - \mathbb{E}\,(X_t X_{t+s}^n) + \mathbb{E}\,(X_t X_{t+s}^n) - \mathbb{E}\,(X_t^n X_{t+s}^n)| \\
&= |\mathbb{E}\,(X_t(X_{t+s} - X_{t+s}^n)) + \mathbb{E}\,(X_{t+s}^n(X_t - X_t^n))| \\
&= |\langle X_t, \underbrace{X_{t+s} - X_{t+s}^n}_{\xrightarrow[n\to\infty]{L^2}0}\rangle_{L^2} + \langle \underbrace{X_{t+s}^n}_{\xrightarrow[n\to\infty]{L^2}0}, \underbrace{X_t - X_t^n}_{\xrightarrow[n\to\infty]{L^2}0}\rangle_{L^2}| \\
&\overset{CS-Ungl.}{\leq} ||X_t||_{L^2} \cdot ||X_{t+s} - X_{t+s}^n||_{L^2} \\
&\quad + \underbrace{||X_{t+s}^n||_{L^2}}_{\leq ||X_{t+s}||_{L^2} < \infty} \cdot \underbrace{||X_t - X_t^n||_{L^2}}_{\xrightarrow[n\to\infty]{L^2}0} \xrightarrow[n\to\infty]{} 0 \\
&\Rightarrow \mathrm{Cov}\,(X_t, X_{t+s}) = \sigma^2 \sum_{i=0}^{\infty} \gamma_i \gamma_{i+|s|}
\end{aligned}
$$

□

2) Prozesse der gleitenden Mittel

Definition 3.2.5

Sei $\alpha_0, \ldots, \alpha_q$ eine endliche Zahlenfolge mit $\alpha_0 = 1,\ \alpha_q \neq 0,\ q \in \{0, 1, 2, \ldots\}$. 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 $\beta_0, \ldots, \beta_p$ eine endliche Zahlenfolge mit $\beta_0 = 1,\ \beta_p \neq 0$ für ein $p \in \{0, 1, 2, \ldots\}$. Der stochastische Prozess $X = \{X_t\}_{t\in\mathbb{Z}}$ heißt *autoregressiver Prozess der Ordnung* p (AR(p)), falls X eine Lösung folgender Differenzengleichung 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 Nullstellen $z_1, \ldots, z_p \in \mathbb{C}$ 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. $z_j \in B_1(0) \ \forall \ j = 1, \ldots, p$.

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

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

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

Ein Beweis befindet sich in „*Löbus - Ökonometrie*“.

4) Autoregressive Prozesse mit gleitendem Mittel

Definition 3.2.7

Für $p, q \in \{0, 1, 2, \ldots\}$ seien $\alpha_0, \ldots, \alpha_q$ bzw. $\beta_0, \ldots, \beta_q$ 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 Differenzengleichung

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

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 $p, q \in \{0, 1, 2, \ldots\}$. 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 $\Delta X = \{\Delta X_t\}_{t\in Z}$ der Zuwächse

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

von X ein ARMA(p, q) - Prozess ist. Allgemeiner, für den rekursiv 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 $X_{t-1}, \ldots, X_{t-p}$ vorhersagbar ist oder dessen Vorhersage Unsicherheiten enthält.

Definition 3.2.9

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

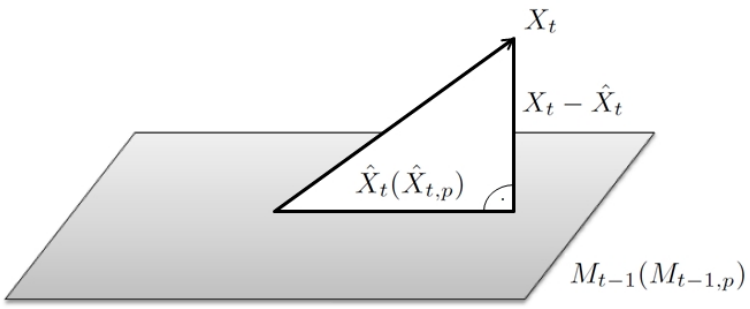
1. Der *Unterraum* M_{t-1}, $t \in \mathbb{Z}$, wird definiert als linearer Abschluss in $L^2(\Omega, \mathcal{F}, P)$ von der Folge $\{X_{t-1}, X_{t-2}, \ldots\}$, die die Vergangenheit von X bis zum Zeitpunkt $t \in \mathbb{Z}$ darstellt, d.h. M_{t-1} enthält alle Linearkombinationen der Zufallsvariablen aus der Folge $\{X_{t-1}, X_{t-2}, \ldots\}$ 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 $\langle X_{t-1}, \ldots, X_{t-p}\rangle$ von Zufallsvariablen $X_{t-1}, \ldots, X_{t-p}$, d.h. $M_{t-1,p}$ besteht aus allen Linearkombinationen von $X_{t-1}, \ldots, X_{t-p}$. $M_{t-1,p}$ 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. Approximiere X_t durch

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

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 $X_{t-1}, X_{t-2}, \ldots$ von X mit unendlichem Zeithorizont:

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

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 $X_{t-1}, \ldots, X_{t-p}$ von X mit endlichem Zeithorizont und $p \in \mathbb{N}$:

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

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 Prozess[4] 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 $\mathbb{E}\,X_t \equiv 0$.

[4]es handelt sich im Folgenden immer um einen stationären Prozess im weiteren Sinne

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 $\hat{X}_t \overset{\text{f.s.}}{=} X_t$ (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

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

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}, $s = 1, \ldots, p$, 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

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

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

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

$R = \{R_t\}_{t\in\mathbb{Z}}$ 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 $X_{t_1}, \ldots, X_{t_q}, (q \leq p)$, eine Basis von $M_{-1,p} = \langle X_{-1}, \ldots, X_{-p}\rangle$. Beachte, dass $t_1, \ldots, t_q < 0$.
   Sei $K(s) = \mathbb{E}\,(X_0 X_s)$ die Kovarianzfunktion des Prozesses X. Es gilt

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

   wobei

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

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 $S_{t_1}, \ldots, S_{t_q}$ eine Basis von $\langle S_{-1}, \ldots, S_{-p}\rangle,\ q \leq p$. Dann gilt

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

   wobei

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

3.2.4 Eigenschaften der Zeitreihen

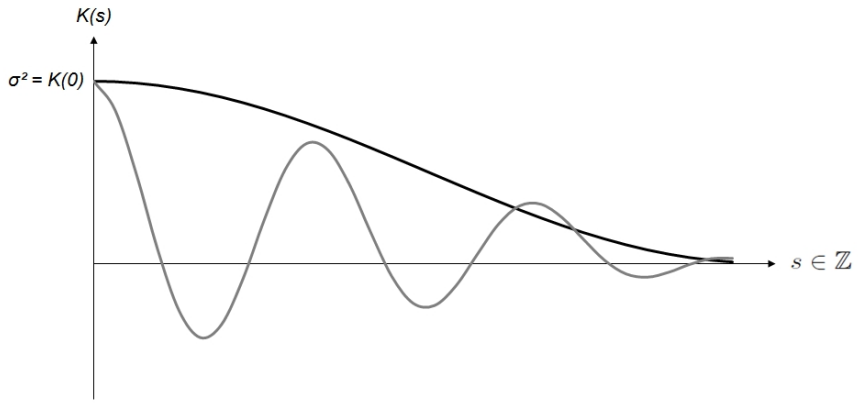
Definition 3.2.12

Sei $X = \{X_t\}_{t\in\mathbb{Z}}$ ein stationärer Prozess mit $\mathbb{E}\, X_t \equiv 0$ und $\sigma^2 = \operatorname{Var} X_t > 0$. 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.

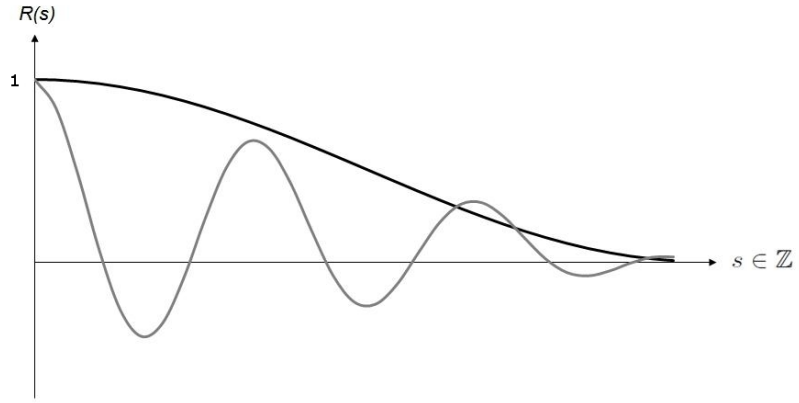
Es gilt: $K(s) = K(-s)$, $\quad R(s) = R(-s)$.

Beweis

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



(Auto-)Kovarianzfunktion



(Auto-)Korrelationsfunktion

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

Beispiele

1) Lineare Prozesse

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

Aus dem Satz 3.2.2 folgt

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

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

$$|K(s)| \overset{CS-Ungleichung}{\leq} \sigma^2 \sqrt{\sum_{j=0}^{\infty} \gamma_{j+|s|}^2 \sum_{j=0}^{\infty} \gamma_j^2} = \sigma^2 \sqrt{\sum_{i=|s|}^{\infty} \gamma_i^2 \sum_{j=0}^{\infty} \gamma_j^2} \xrightarrow[s \to +\infty]{} 0 \ ,$$

weil

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

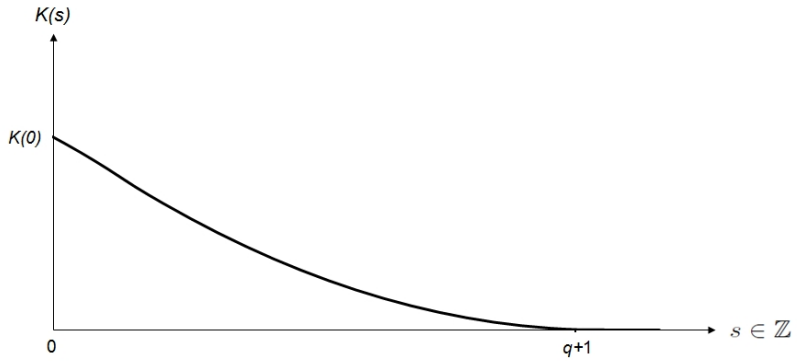
2) MA(q)-Prozesse

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

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

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

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



Satz 3.2.6

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

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

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

(ohne Beweis)

[4] Ungleichung von Cauchy-Schwarz

3) Stationäre AR(p)-Prozesse

Sie werden definiert als Lösung des Gleichungssystems

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

Unter der Bedingung des Satzes 3.2.3 $\exists \{\gamma_j\}_{j=0}^{\infty}$ 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 $\{\beta_0, \ldots, \beta_p\}$ gegeben. Im Folgenden geben wir ein Verfahren an, das die Werte $K(s)$ bzw. $R(s)$ als Funktion von $\{\beta_0, \ldots, \beta_p\}$ liefert.

Satz 3.2.7 (Yule-Walker-Gleichungen)

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

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

Zusätzlich gilt:

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

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

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

z.B.

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

Beweis

Aus der Definition von AR(p) gilt

$$\delta_t = \beta_0 X_t + \ldots + \beta_p X_{t-p} \ . \quad (*)$$

Aus der linearen Darstellung von X erhalten wir

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

Durch Ausmultiplizieren der beiden Seiten der letzten Gleichung mit $(*)$ und Berechnung des Erwartungswertes gilt

$$0 = \sum_{j=0}^{\infty} \underbrace{\mathbb{E}\,(\delta_t \delta_{-j})}_{\substack{=0 \\ \text{wegen der} \\ \text{Unkorreliert-} \\ \text{heit von } \{\delta_t\}}} \gamma_j = \beta_0 \underbrace{\mathbb{E}\,(X_0 X_t)}_{K(t)} + \ldots + \beta_p \underbrace{\mathbb{E}\,(X_0 X_{t-p})}_{=K(t-p)}, \quad t > 0\ . \tag{3.2.6}$$

Durch die Division mit $K(0)$ erhalten wir die Yule-Walker-Gleichung:

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

Für $t = 1, \ldots, p-1$

$$\left.\begin{array}{c} \sum\limits_{j=0}^{p} \beta_j R(1-j) = 0 \\ \vdots \\ \sum\limits_{j=0}^{p} \beta_j R(p-1-j) = 0 \end{array}\right| \Rightarrow \text{ Ergebnis (3.2.4)},$$

wobei $R(0) = 1$ und $R(-s) = R(s)$, $s \in \mathbb{Z}$ verwendet werden.

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

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

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

□

4) Stationäre ARMA(p,q)-Prozesse

Der Ausdruck für $K(s)$ bzw. $R(s)$ als Funktion von $\{\alpha_0, \ldots, \alpha_q\}$ und $\{\beta_0, \ldots, \beta_p\}$ 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 $F : [-\pi; \pi] \to [0; K(0)]$ mit den Eigenschaften $F(-\pi) = 0,\ F(\pi) = K(0) = \operatorname{Var} X_t,$ so dass

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

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. f ist eine symmetrische Funktion:

$$f(-s) = f(s) \ , \quad s \in [-\pi;\pi] \ .$$

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 die Spektraldichte

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

2. MA(q)-Prozesse

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

$\Rightarrow X$ besitzt die 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 $\sigma^2 = \operatorname{Var}\delta_j,\ \{\beta_0, \ldots, \beta_p\}$ 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 $\sigma^2 = \operatorname{Var}\delta_j,\ \{\alpha_0, \ldots, \alpha_q\}$ und $\{\beta_0, \ldots, \beta_p\}$ die Koeffizienten aus der Definition 3.2.7 von X sind.

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. Unter Berücksichtigung der Tatsache, dass 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 unbekannten Koeffizieten 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 vor Parametern abhängen, damit der Bezug zu ökonometrischen Anwendungen nicht verloren geht.

3.3.1 Schätzung der Kovarianzfunktion

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

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

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

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

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

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

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

3.3.2 Parameterschätzung bei AR(p)-Prozessen

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

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

wobei $\beta_0 = 1,\ \beta_p \neq 0,\ \delta_j \sim \mathcal{N}(0,\sigma^2),\ \{\delta_j\}$ unkorreliert, $\sigma^2 > 0$. Der Prozess X ist eindeutig durch $\beta_1, \ldots, \beta_p$ und σ^2 festgelegt. Wir stellen zwei Methoden zur Schätzung von $\beta_1, \ldots, \beta_p$ 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 $X_0 = \ldots = X_{1-p} = 0$ und die Anzahl n der beobachteten Größen $X_1, \ldots, X_n$ der Ungleichung $n > p$ genügt, dann kommt man durch mehrfache Anwendung von (3.3.2) für $t = 1, \ldots, n$ zu folgendem linearen Gleichungssystem:

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

wobei

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

$$\beta = (\beta_1, \ldots, \beta_p)^\top \ , \quad \delta = (\delta_1, \ldots, \delta_n)^\top \ .$$

Nehmen wir an, dass $rg(A) = p$.
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 Nullstellen 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:

$$\hat{\beta}_n \xrightarrow[n\to\infty]{\text{f.s.}} \beta \ , \quad \hat{\sigma}_n^2 \xrightarrow[n\to\infty]{\text{f.s.}} \sigma^2.$$

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)

$$R(t) = -\sum_{\substack{j=1 \\ j\neq t}}^{p} \beta_j R(t-j) \ ,$$

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 Komponente $X = \{X_t\}_{t\in\mathbb{Z}}$ 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 3.2.4

3.3.3 Parameterschätzung bei ARMA(p,q)-Prozessen

Sei $X = \{X_t, t \in \mathbb{Z}\}$ ein ARMA(p,q)-Prozess, p,$q \in \mathbb{N}_0$, d.h.

$$\sum_{i=0}^{q} \alpha_i \varepsilon_{t-i} = \sum_{j=0}^{p} \beta_j x_{t-j} , \quad t \in \mathbb{Z} , \quad p+q \geq 1 ,$$

wobei $\{\varepsilon_t\}$ eine Folge von unabhängigen Zufallsvariablen mit $\mathbb{E}\,\varepsilon_t = 0$, $\mathbb{E}\,\varepsilon_t^2 = \sigma^2$, $t \in \mathbb{Z}$, $t_1, \dots, t_q \in \mathbb{R}$, $t_q \neq 0$ ist und $\beta_1, \dots, \beta_p \in \mathbb{R}$, $\beta_p \neq 0$.

Alle Nullstellen der Polynome

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

$$Q(z) = \sum_{k=0}^{q} \alpha_k z^{q-k} , \quad z \in \mathbb{C},$$

liegen in $\{z \in \mathbb{C}: |z| < 1\}$.

Sei für jedes $r \in \mathbb{N}$
$Q_r = \big\{(a,b)$: $a = (a_1, \dots, a_q)^\top \in \mathbb{R}^q$, $b = (b_1, \dots, b_p)^\top \in \mathbb{R}^p \, |a_1|, \dots, |a_q|, |b_1|, \dots, |b_p| < r\big\}$ und die Polynome $\sum_{k=0}^{q} a_k z^{q-k}$ und $\sum_{k=0}^{p} b_k z^{p-k}$,$z \in \mathbb{C}$, besitzen keine gemeinsamen Nullstellen.

Sei $\Pi_j \in \mathbb{R}$ mit $|\Pi_j| \leq re^{-\frac{j}{r}}$, $j \geq 0$, wobei $\Pi_j = \Pi_j(a,b)$:

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

wobei $Y = \{Y_t, t \in \mathbb{Z}\}$ ein ARMA(p,q)-Prozess ist, der mit Hilfe der Koeffizienten $a_1,\ldots,a_q$ und $b_1,\ldots,b_p$ definiert ist.
Sei $X = \{X_t, \ t \in \mathbb{Z}\}$ so, dass $(\alpha,\beta) \in Q_r$ für ein $r \in \mathbb{N}$. Unter diesen Voraussetzungen ist unser ARMA(p, q)-Prozess X linear und stabil.

Seien die Beobachtungen von X, nämlich $X_1, \ldots, X_n$, gegeben, $n > p + q$.

Ziel:

Schätze $\alpha = (\alpha_1, \ldots, \alpha_q)^\top$ und $\beta = (\beta_1, \ldots, \beta_p)^\top$ aus den Daten $X_1, \ldots, X_n$.
Wir wissen:

$$\sum_{i=0}^{q} \alpha_i \varepsilon_{t-i} = \sum_{j=0}^{p} \beta_j X_{t-j} \ , \quad \varepsilon_t \text{ nicht beobachtbar.}$$

Schätzungsansatz:

Berechne ε_j und schätze daraus α, β.
Setze $\varepsilon_0 = \varepsilon_{-1} = \ldots = \varepsilon_{1-p} = 0$ und $X_0 = X_{-1} = \ldots = X_{1-p} = 0$.

$$\begin{aligned}
\varepsilon_1 &= X_1 - \underbrace{\sum_{i=1}^{q} \alpha_i \varepsilon_{1-i}}_{0} + \underbrace{\sum_{j=1}^{p} \beta_j X_{t-j}}_{0} = X_1 \\
\varepsilon_2 &= X_2 + \beta_1 X_1 - \alpha_1 \varepsilon_1 \\
\varepsilon_3 &= X_3 + \beta_1 X_2 + \beta_2 X_1 - \alpha_1 \varepsilon_2 - \alpha_2 \varepsilon_1 \\
&\vdots \\
\varepsilon_t &= X_t + \sum_{j=1}^{min\{t-1,p\}} \beta_j X_{t-j} - \sum_{i=1}^{q} \alpha_i \varepsilon_{t-1} \ , \quad t = 1, \ldots, n \ .
\end{aligned}$$

Dies ist ein Gleichungssystem, das $\varepsilon_j, j = 1, \ldots, n$ rekursiv durch $\alpha_1, \ldots, \alpha_q, \beta_1, \ldots, \beta_p$ definiert.
Man berechnet $(\alpha, \beta) \in Q_r$, indem man

$$\sum_{i=1}^{n} \varepsilon_i^2 \longrightarrow \min_{\alpha,\beta} \ . \tag{3.3.3}$$

Als Begründung sei angemerkt, dass

$$\hat{\sigma}^2 = \frac{1}{n-p-q} \sum_{i=1}^{n} \hat{\varepsilon}_i^2$$

ein Schätzer für $\operatorname{Var} \varepsilon_i = \sigma^2$ ist.
Diese Methode ist als bedingte Methode der kleinsten Quadrate bekannt. Für die Lösung der Gleichung (3.3.3) muss man numerische Optimierungsmethoden einsetzen, d.h. sie kann nicht explizit angegeben werden. Dennoch sind $\hat{\alpha}, \hat{\beta}$ konsistente Schätzer für α und β, wie folgender Satz zeigt:

Satz 3.3.1

Sei $\mathbb{E}\, \varepsilon_t^4 < \infty$ unabhängig von $t \in \mathbb{Z}$ ab. Unter den obigen Voraussetzungen gilt:

$$\hat{\alpha} \xrightarrow[n\to\infty]{\text{f.s.}} \alpha \ , \quad \hat{\beta} \xrightarrow[n\to\infty]{\text{f.s.}} \beta \ , \tag{3.3.4}$$

d.h. beide Schätzer sind stark konsistent.
Ohne Beweis.

3.3.4 Schätzung der Spektraldichte

Sei $X = \{X_t, t \in \mathbb{Z}\}$ ein stationärer Prozess mit $\mathbb{E}\,[X_t] = 0, t \in \mathbb{Z}$, und einer Kovarianzfunktion $K : \mathbb{Z} \to \mathbb{R}, K(s) = \mathbb{E}\,[X_0 X_s]$, $s \in \mathbb{Z}$.
Wir nehmen an, dass die Spektraldichte von X existiert, d.h.

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

Wir wissen, dass K mit $\hat{K}(s) = \int_{-\pi}^{\pi} e^{its} \hat{f}(t) dt$ geschätzt werden kann, falls

$$\hat{f} : [-\pi, \pi] \to \mathbb{R}$$

eine geeignete Schätzung von f ist. Deshalb werden wir jetzt versuchen, unterschiedliche Schätzansätze für f vorzuschlagen.

Definition 3.3.1

Die Funktion $I_n(s) = \frac{1}{2\pi n} \left| \sum e^{-its} X_t \right|^2$,$s \in [-\pi, \pi]$, heißt Periodogramm des Prozesses $X = \{X_t, t \in \mathbb{Z}\}$.

Satz 3.3.2 Sei f stetig. Das Periodogramm ist dann ein asymptotisch erwartungstreuer Schätzer für f, d.h. $\mathbb{E}\,(I_n(s)) \xrightarrow[n\to\infty]{} f(s)$, $s \in [-\pi, \pi]$

Beweis:

Es gilt $I_n(s) \geq 0, \quad s \in [-\pi, \pi]$, und $I_n \in L^1(\Omega, \mathcal{F}, P)$, $\mathbb{E}\,(I_n(s)) < \infty$.

$$\begin{aligned}
\mathbb{E}\,(I_n(\lambda)) &\leq \frac{1}{2\pi n}\mathbb{E}\left(\left|\sum_{s,t=1}^{n} e^{-i\lambda s - i\lambda t} X_s X_t\right|\right) \\
&\leq \frac{1}{2\pi n}\sum_{s,t=1}^{n} \mathbb{E}\,(|X_s X_t|) \\
&\overset{CS-Ungleichung}{\leq} \frac{1}{2\pi n}\sum_{s,t=1}^{n} \sqrt{\mathbb{E}\,(X_s^2)}\sqrt{\mathbb{E}\,(X_t^2)} < \infty \\
E\,(I_n(\lambda)) &= \frac{1}{2\pi n}\mathbb{E}\left(\left|\sum_{t=1}^{n} e^{-it\lambda} X_t\right|^2\right) = \frac{1}{2\pi n}\sum_{s,t=1}^{n} e^{-i(t-s)\lambda}\underbrace{\mathbb{E}\,(X_s X_t)}_{K(s-t)} \\
&= \frac{1}{2\pi n}\sum_{s,t=1}^{n} e^{-i(t-s)\lambda}\int_{-\pi}^{\pi} e^{i(s-t)x} f(x)dx \\
&= \frac{1}{2\pi n}\int_{-\pi}^{\pi}\left|\sum_{t=1}^{n} e^{it(x-\lambda)}\right|^2 f(x)dx \xrightarrow[n\to\infty]{} f(x)
\end{aligned}$$

$$\text{falls } \frac{1}{2\pi n}\left|\sum_{t=1}^{n} e^{it(x-\lambda)}\right|^2 \xrightarrow[n\to\infty]{} \delta_\lambda(x) = \begin{cases} 1\,, & x = \lambda \\ 0\,, & \text{sonst.} \end{cases}$$

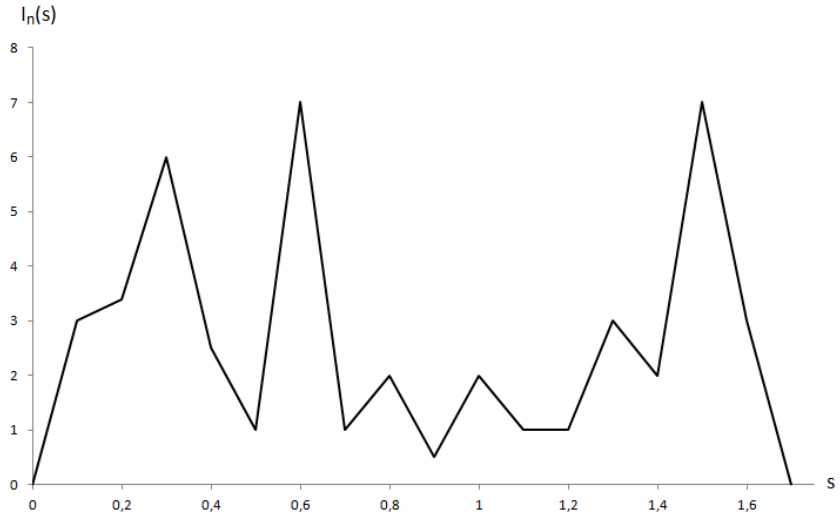
□

I_n ist allerdings kein konsistenter Schätzer für $f \in [-\pi, \pi]$, d.h.

$$I_n(s) \overset{P}{\not\longrightarrow}_{n\to\infty} f(s) \text{ für } s \in [-\pi, \pi].$$

Es gilt u.a. $\operatorname{Var} I_n \not\longrightarrow_{n\to\infty} 0$. Deswegen hat die Grafik von I_n typischerweise mehrere „Zacken“, d.h. sie zeigt große Oszilationen auf.

Beispiel 3.3.1



Ein Periodogramm $I_n(s)$

Sei nun $b \in L^1([-\pi,\pi])$, d.h. $\int_{-\pi}^{\pi} |b(s)|ds < \infty$.

Definition 3.3.2

Die Zufallsvariable $I_{n,b} = \int_{-\pi}^{\pi} I_n(s)b(s)ds$ heißt integriertes Periodogramm.

Da $f \in C[\pi,\pi]$, wobei $C[\pi,\pi]$ die Menge der stetigen Funktionen auf $[\pi,\pi]$ ist, ist f beschränkt auf $[-\pi,\pi]$, d.h. $\exists M \in (0,\infty)$:

$$0 \leq f(s) \leq M \ , \quad \forall s \in [\pi,\pi] \ .$$

Zeigen wir, dass $I_{n,b} \in L^1(\Omega,\mathcal{F},P)$, d.h. $\mathbb{E}\,|I_{n,b}| < \infty \ \forall n \in \mathbb{N}$ und alle $b \in L^1([-\pi,\pi])$. tatsächlich gilt

$$\begin{aligned}
\mathbb{E}\,|I_{n,b}| &= \mathbb{E}\left|\int_{-\pi}^{\pi} I_n(s)b(s)ds\right| \\
&\overset{Satz von Fubini}{\leq} \int_{-\pi}^{\pi} \mathbb{E}\,|I_n(s)|\,|b(s)|\,ds \\
&\leq M \int_{-\pi}^{\pi} |b(s)|\,ds \\
&< \infty \ .
\end{aligned}$$

Für das Periodogramm gilt:

$$\begin{aligned}\mathbb{E}\left[I_n(\lambda)\right] &= \frac{1}{2\pi n}\int_{-\pi}^{\pi} \underbrace{\left|\sum_{t=1}^{n} e^{it(s-\lambda)}\right|^2}_{\leq\left(\sum_{i=1}^{n}|e^{it(s-\lambda)}|\right)^2=n^2} f(s)ds \\ &\leq n^2 \cdot 2\pi \cdot \frac{M}{2\pi n} \\ &= n \cdot M\end{aligned}$$

Eine bessere Abschätzung bekommt man durch die Tatsache, dass

$$\frac{1}{2\pi n}\int_{-\pi}^{\pi}\left|\sum_{t=1}^{n} e^{it(s-\lambda)}\right|^2 ds = 1$$

und damit

$$\mathbb{E}\left[I_n(\lambda)\right] = M \cdot \frac{1}{2\pi n}\int_{-\pi}^{\pi}\left|\sum_{t=1}^{n} e^{it(s-\lambda)}\right|^2 ds = \mu\ .$$

Lemma 3.3.3

Unter den obigen Voraussetzungen ist $I_{n,b}$ ein asymptotisch erwartungstreuer Schätzer für $\int_{-\pi}^{\pi} f(s)b(s)ds$, d.h. $\mathbb{E}\left[I_{n,b}\right] \xrightarrow[n\to\infty]{} \int_{-\pi}^{\pi} f(s)b(s)ds$.

Beweis:

$$\begin{aligned}\lim_{n\to\infty}\mathbb{E}\left[I_{n,b}\right] &= \lim_{n\to\infty}\mathbb{E}\left[\int_{-\pi}^{\pi} I_n(s)b(s)ds\right] \\ &= \lim_{n\to\infty}\int_{-\pi}^{\pi}\mathbb{E}\left[I_n(s)\right]b(s)ds \\ &\overset{\text{Satz von Lebesgue über die majorisierte Konvergenz}}{=} \int_{-\pi}^{\pi}\lim_{n\to\infty}\mathbb{E}\left[I_n(s)\right]b(s)ds \\ &= \int_{-\pi}^{\pi} f(s)b(s)ds,\end{aligned}$$

weil $\mathbb{E}\left[I_n(s)\right] \xrightarrow[n\to\infty]{} f(s)\ \ \forall s \in [-\pi,\pi]$ (siehe Satz 3.3.2).

□

Lemma 3.3.4

Unter den obigen Voraussetzungen gilt

$$I_{n,b} = \sum_{j=-(n-1)}^{n-1} w_j \hat{K}(j) \ ,$$

wobei $\hat{K}(s) = \frac{1}{n}\sum_{t=n}^{n-2} X_t X_{t+s}$ ein Schätzer für die Kovarianzfunktion K ist und $w_j,\ j \in \mathbb{Z}$ Gewichte sind mit

$$w_j = \frac{1}{2\pi n} \int_{-\pi}^{\pi} e^{-ijs} b(s) ds \ , \quad j \in \mathbb{Z} \ .$$

Beweis:

$$\begin{aligned}
\text{Früher: } I_n(\lambda) &= \frac{1}{2\pi n} \sum_{s,t=1}^{n} e^{-i(t-s)\lambda} X_s X_t \\
&= \frac{1}{2\pi} \underbrace{\frac{1}{n}\left(\sum_{t=1}^{n} X_t^2\right)}_{\hat{K}(0)} + \frac{1}{2\pi n}\left(\sum_{\substack{s,t=1 \\ s>t}}^{n} e^{-i(t-s)\lambda} X_s X_t + \sum_{\substack{s,t=1 \\ s<t}}^{n} e^{-i(t-s)\lambda} X_s X_t\right) \\
&= \frac{1}{2\pi}\left(\hat{K}(0) + \frac{1}{n}\sum_{k=1}^{n-1}\sum_{s=1}^{n-k} e^{-ik\lambda} X_s X_{s+k} + \frac{1}{n}\sum_{k=1}^{n-1}\sum_{s=1}^{n-k} e^{-ik\lambda} X_t X_{t+k}\right) \\
&= \frac{1}{2\pi}\left(\hat{K}(0) + \sum_{k=1}^{n-1} e^{-ik\lambda}\hat{K}(k) + \sum_{k=1}^{n-1} e^{ik\lambda}\hat{K}(-k)\right) \\
&= \frac{1}{2\pi}\left(\sum_{k=-(n-1)}^{n-1} e^{-ik\lambda}\hat{K}(k)\right) \\
\Rightarrow I_{n,b} &= \int_{-\pi}^{\pi} I_n(s) b(s) ds \\
&= \int_{-\pi}^{\pi} \frac{1}{2\pi} \sum_{j=-(n-1)}^{n-1} e^{-ijs}\hat{K}(s) b(s) ds \\
&= \sum_{j=-(n-1)}^{n-1} \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-ijs} b(s) ds \cdot \hat{K}(s).
\end{aligned}$$

□

Der Vorteil von $I_{n,b}$ liegt gegenüber I_n darin, dass $I_{n,b}$ ein konsistenter Schätzer für $\int_{-\pi}^{\pi} f(s)b(s)ds$ ist.

Satz 3.3.3

Sei $X = \{X_t, t \in \mathbb{Z}\}$ ein linearer Prozess, d.h. $X_t = \sum_{j=0}^{\infty} \gamma_j \varepsilon_{t-j}$, wobei $\varepsilon_t, t \in \mathbb{Z}$ eine

Folge von unabhängigen Zufallsvariablen ε_t ist mit $\mathbb{E}[\varepsilon_t] = 0$, und $\mathbb{E}\,\varepsilon_t^2 = \sigma^2$, $t \in \mathbb{Z}$, und $\mathbb{E}\,\varepsilon^4 < \infty$ hängt nicht von t ab. Weiterhin gelte $|\gamma_j| \leq C \cdot e^{-aj},\ \forall j \in \mathbb{N}_0$, und für ein $C > 0$, $a > 0$.

Sei $f : [-\pi, \pi] \to \mathbb{R}$ die Spektraldichte von X, $b \in L^1([-\pi, \pi])$ mit $b(s) = g_1(s) + g_2(s)$, wobei g_1 diff'bar auf $[-\pi, \pi]$, $g_1' \in L^2([-\pi, \pi])$, $g_2(s) = \sum_{j=1}^m c_j \cdot I(s \leq d_j),\ s \in [-\pi, \pi]$ für Konstanten $c_1, ..., c_m \in \mathbb{R},\ d_1, ..., d_m \in [-\pi, \pi]$. Dann ist $I_{n,b}$ ein schwach konsistenter Schätzer für $\int_{-\pi}^{\pi} f(s) \cdot b(s) ds$, d.h. $I_{n,b} \xrightarrow[n\to\infty]{P} \int_{-\pi}^{\pi} f(s) \cdot b(s) ds$.

Ohne Beweis.

Bemerkung 3.3.2

1. Die Konsistenzeigenschaften von $I_{n,b}$ werden besser im Vergleich zu denen von I_n, weil man durch die Integration mit der Funktion b die „Unregelmäßigkeiten“ des Periodogramms I_n glättet.
2. Falls $\varepsilon_t \sim N(0, \sigma^2), t \in \mathbb{Z}$, d.h. der lineare Prozess x aus dem Satz 3.3.3 ist Gauß´sch, dann kann gezeigt werden, dass $\lim_{n\to\infty}(n \cdot Var(I_{n,b})) = 4\pi \int_{-\pi}^{\pi} f^2(s) b^2(s) ds\ \forall f, b$ wie in Satz 3.3.3, so dass zusätzlich b beschränkt auf $[-\pi, \pi]$ und symmetrisch ist (d.h. $b(-s) = b(s)\ \forall s \in [-\pi, \pi]$). Daraus folgt, dass $\mathrm{Var}(I_{n,b}) \underset{n\to\infty}{=} O(\frac{1}{n})$.

Beispiele:

1. Schätzung der Kovarianz:

   Falls $b(s) = e^{ist}$ für $t \in \mathbb{Z}\ ,\quad s \in [-\pi, \pi]$, dann ist

   $$\begin{aligned}\hat{k} &= \int_{\pi}^{-\pi} I_n(s) b(s) ds \\ &= \int_{\pi}^{-\pi} I_n(s) \cos(st) ds\ ,\quad t \in \mathbb{Z},\end{aligned}$$

   ein konsistenter Schätzer für $k(t),\ n \to \infty,\ t = -(n-1), ..., (n-1)$.

2. Schätzung der Spektralfunktion:

$$F(\lambda) = \int_{-\pi}^{\lambda} f(s)ds \ , \quad \rightarrow \text{Spektralfunktion}$$

$$= \int_{\pi}^{-\pi} f(s) \cdot \underbrace{I(s \in [-\pi, \lambda])}_{b(s)} ds \ , \quad \lambda \in [-\pi, \pi] \ .$$

$$K(t) = \int_{\pi}^{-\pi} e^{ist} dF(s)$$

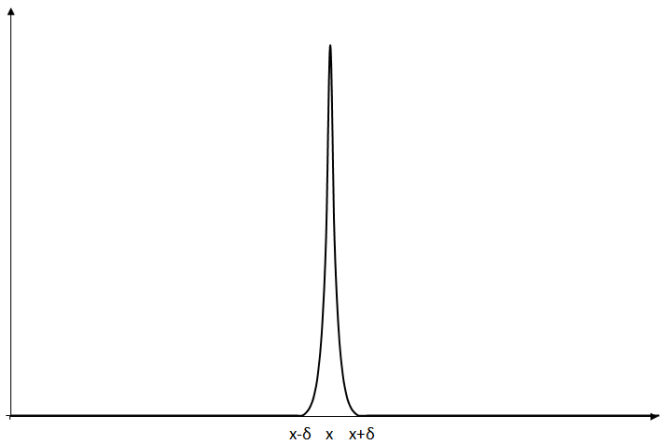
$\Rightarrow$ Der Schätzer $\hat{F}_n(\lambda) = \int_{-\pi}^{\lambda} I_n(s)ds$ ist ein schwach konsistenter Schätzer für $F(\lambda)$, $\lambda \in [-\pi, \pi]$.

3.3.5 Beispiele für die Schätzung der Spektraldichte

Sei $X = \{X_t : t \in \mathbb{Z}\}$ eine zentrierte Zeitreihe mit der Spektraldichte f, d.h. $K(s,t) = \mathbb{E}[X(s), X(t)] = \int_{-\pi}^{\pi} e^{ix(s-t)} f(x)dx$. Sei $I_n(x)$ das Periodogramm, aufgebaut auf den beobachteten Werten $x_1, ..., x_n$, ein erwartungstreuer nicht konsistenter Schätzer für f.

Daher ist das integrierte Periodogramm: $I_{n,b}(x) = \int_{-\pi}^{\pi} I_n(y) \cdot b(x-y)dy$, wobei die Funktion $b : \mathbb{R} \rightarrow \mathbb{R}_+$ integrierbar auf $[-\pi, \pi]$ und periodisch mit Periode π und symmetrisch ist: $b(x) = b(-x) \forall b \in \mathbb{R}$. $I_{n,b}$ ist ein konsistenter Schätzer für $\int_{-\pi}^{\pi} f(y)b(x-y)dy$.

Für welche Wahl von b gilt: $\int_{-\pi}^{\pi} f(y)b(x-y)dy \approx f(x)$?



Wähle b so, dass sie fast Null ist, außer für $|y| \leq \delta$ und $\int_{-\infty}^{\infty} b(y)dy = 1$.

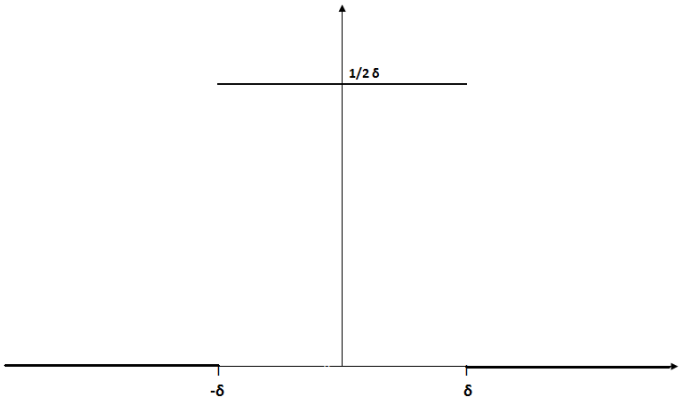
Bezeichnung: $b(y) = b_\delta(y)$. Es ist klar, dass

$$b_\delta(\cdot) \underset{\delta\to 0}{\to} \delta_0(\cdot) \Longrightarrow \int_{-\pi}^{\pi} f(y)b(x-y)dy \underset{\delta\to 0}{\to} f(x)\ , \quad \forall x \in \mathbb{R}.$$

1. Daniel-Schätzer:

$$b_\delta(y) = \begin{cases} \frac{1}{2\delta}, & |y| \le \delta \\ 0, & y \in [-\pi,\pi]\ , \quad \text{Fortsetzung auf } \mathbb{R} \text{ periodisch} \end{cases}$$

$$\Longrightarrow I_{n,b}(x) = \frac{1}{2\delta}\int_{x-\delta}^{x+\delta} I_n(y)by \underset{\delta\to 0, n\to\infty}{\overset{\text{f.s.}}{\longrightarrow}} f(x)\ .$$



Der Daniel-Schätzer

2. Parzen-Schätzer:

   Sei r gerade, $r < n$ (empfohlen $\frac{n}{6} \le r \le \frac{n}{5}$),

$$b(y) = \frac{1}{\pi r^3}\left(\frac{\sin^2(ry/2) - 4\sin^2(ry/4)}{\sin^2(y/2)} + \frac{12\sin^2(ry/4) - 3\sin^2(ry/2)}{2\sin^4(y/2)}\right),$$

   $y \in (-\pi,\pi)$, für $y \in \mathbb{R}$ wird b periodisch fortgesetzt. Dieser Schätzer besitzt eine kleine Varianz $\operatorname{Var} I_{n,b}$ und einen kleinen Bias $|\mathbb{E}\,(I_{n,b}(x) - f(x))|$.

3. Barlett-Schätzer:

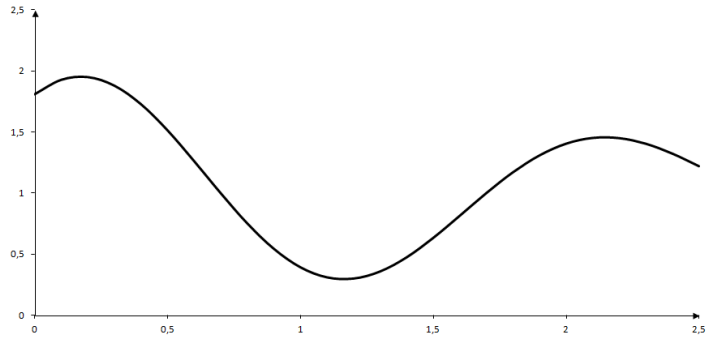
   Für ein $r < n$ definiere

$$b(y) = \frac{n}{2\pi}\left(\frac{\sin((2r+1)y/2)}{\sin(y/2)} - (n-r)\sum_{t=-r}^{r}\frac{\cos(ty)}{n-|t|}\right),$$

   $y \in (-\pi,\pi)$, für $y \in \mathbb{R}$ wird b periodisch fortgesetzt.

3.4 Zeitreihen mit stationärem Anteil

Sei $Z = \{Z_t,\ t \in \mathbb{Z}\}$ eine Zeitreihe, $Z_t = S_t + X_t \quad t \in \mathbb{Z}$, wobei $s_t = \sum_{k=1}^{n} b_k \cos(\mu_k t + \varphi_k)$ $t \in \mathbb{Z}$ der saisonale Anteil ist. Es sind $\underbrace{b_k \in \mathbb{R}}_{\text{Amplitude}}$, $\underbrace{[\mu_k \in [0,\pi]}_{Frequenz}$, $\underbrace{\varphi_k \in (-\pi/2,\pi/2)}_{Phasenverschiebung}$ deterministisch, aber unbekannt.



Barlett-Schätzung der Spektraldichte im Beispiel der Umsatzdaten

$X = \{X_t, t \in \mathbb{Z}\}$ ist ein zufälliger Schwankungsanteil, der ein statischer zentrierter Prozess ist (ARMA(p,q)).

Zielstellung: Schätze n, b_k, μ_k, φ_k aus den Beobachtungen $Z_1, \ldots, Z_n$.

Um effizienter mit Parametern von S_t umgehen zu können, wählen wir die komplexe Darstellung von S_t als

$$S_t = \sum_{k=1}^{v} a_k e^{i\lambda_k t} \ , \quad t \in \mathbb{Z},$$

wobei $a_k \in \mathbb{C}, \lambda_k \in \mathbb{R}$ auf folgende Weise mit b_k, μ_k, φ_k zusammenhängen ($v = 2u + 1$):

$$\begin{array}{ll}
a_1 = \frac{b_u}{2} e^{-i\varphi_u} & \lambda_1 = -\mu_u \\
a_2 = \frac{b_{u-1}}{2} e^{-i\varphi_{u-1}} & \lambda_2 = -\mu_{u-1} \\
\vdots & \vdots \\
a_u = \frac{b_1}{2} e^{-i\varphi_1} & \lambda_u = -\mu_1 \\
a_{u+1} = 0 & \lambda_{u+1} = 0 \\
a_{u+2} = \frac{b_1}{2} e^{i\varphi_1} & \lambda_{u+2} = \mu_1 \\
\vdots & \vdots \\
a_v = \frac{b_u}{2} e^{i\varphi_u} & \lambda_{2u+1} = \mu_u
\end{array}$$

Dies gilt, weil

$$\begin{aligned}
S_t &= \sum_{k=1}^{u} b_k \frac{e^{i(\mu_k t + \varphi_k)} + e^{-i(\mu_k t + \varphi_k)}}{2} \\
&= \sum_{k=1}^{u} \underbrace{\frac{b_k}{2} e^{i\varphi_k}}_{a_k} \underbrace{e^{i\mu_k t}}_{e^{i\lambda_k}} + \sum_{k=1}^{u} \underbrace{\frac{b_k}{2} e^{-i\varphi_k}}_{a_k} e^{-i\mu_k t} \\
&= \sum_{k=1}^{2u+1} a_k e^{i\lambda_k t} \ , \quad \forall t \in \mathbb{Z}.
\end{aligned}$$

Sei x wie oben: $\sum_{j=0}^{p} \beta_j x_{t-j} = \sum_{i=0}^{q} \alpha_i \varepsilon_{t-i}$, $t \in \mathbb{Z}$, $\alpha_0 = \beta_0 = 1$ und $p(z) = \sum_{k=0}^{p} \beta_j z^{p-j}$ hat Nullstellen, die ausschließlich in $\{z \in \mathbb{C} : |z| < 1\}$ liegen. Es gilt $\mathbb{E}\,\varepsilon_i \equiv 0$, $\mathbb{E}\,(\varepsilon_i \varepsilon_j) = \sigma^2 \delta_{ij}, \sigma^2 > 0$.

Sei $I_n(x)$ das Periodogramm von Z:

$$\begin{aligned} I_n(x) &= \frac{1}{2\pi n}\left|\sum_{t=1}^{n} e^{-itx} z_t\right|^2, \quad x \in \mathbb{R} \\ &= \frac{1}{2\pi n}\left|\sum_{t=1}^{n} e^{-itx} S_t + \underbrace{\sum_{t=1}^{n} e^{-itx} X_t}_{\text{im folgenden als } \ldots \text{ bezeichnet}}\right|^2 \\ &= \frac{1}{2\pi n}\left|\sum_{t=1}^{n} e^{-itx} \sum_{k=1}^{v} a_k e^{i\lambda_k t} + \ldots\right|^2 \\ &= \frac{1}{2\pi n}\left|\sum_{k=1}^{v} a_k \underbrace{\sum_{t=1}^{n} e^{it(\lambda_k - x)}}_{(*)} + \ldots\right|^2 \end{aligned}$$

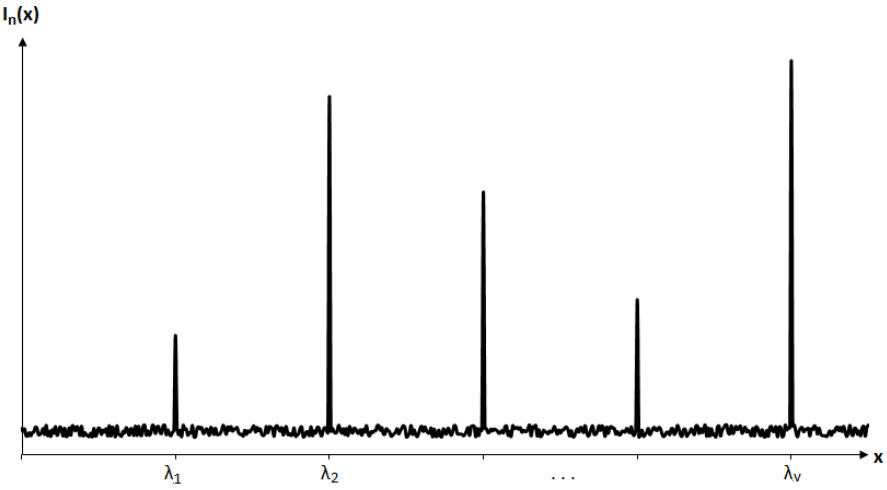
wobei

$$(*) = \begin{cases} \frac{e^{i(n+1)(\lambda_k - x)} - e^{i(\lambda_k - x)}}{i(\lambda_k - x)}, & \text{wenn } \lambda_k \neq x, \\ n, & \text{wenn } \lambda_k = x. \end{cases}$$

$$\Rightarrow I_n(x) = \begin{cases} \frac{1}{2\pi}\left|\frac{1}{\sqrt{n}} \sum_{k=1}^{v} a_k \frac{e^{i(n+1)(\lambda_k - x)} - e^{i(\lambda_k - x)}}{i(\lambda_k - x)} + \right. \\ \qquad \left. \frac{1}{\sqrt{n}} \sum_{t=1}^{n} e^{-itx} X_t\right|^2, & \text{wenn } x \in \mathbb{R} \setminus \{\lambda_1, \ldots, \lambda_n\} \\ \frac{1}{2\pi}\left|\sqrt{n} a_j + \frac{1}{\sqrt{n}} \sum_{k=1, k\neq j}^{v} a_k \frac{e^{i(n+1)(\lambda_k - x)} - e^{i(\lambda_k - x)}}{i(\lambda_k - x)} + \right. \\ \qquad \left. \frac{1}{\sqrt{n}} \sum_{t=1}^{n} e^{-itx} X_t\right|^2, & \text{sonst.} \end{cases}$$

Es gilt

$$\mathbb{E}\left(I_n(x)\right) \begin{cases} \leq c \ \forall n, & x \in \mathbb{R} \setminus \{\lambda_1, \ldots, \lambda_n\}, \text{ wobei } c > 0 \\ \approx n \frac{|a_j|^2}{2\pi}, & x = \lambda_j \text{ für ein } j \in \{1, \ldots, v\} \end{cases}$$



Fazit:

Um u schätzen zu können, muss man die Anzahl der Peaks von I_n bei großen n zählen. Aus I_n lassen sich auch die Werte $|a_j|$ ablesen als $\sqrt{\frac{2\pi}{n}} \cdot$ Höhe des Peaks bei λ_j.
Was ist mit der Schätzung von a_j, λ_j? und daraus folgend b_k, μ_k, φ_k?
Zunächst muss verifiziert werden, ob die Zeitreihe Z einen wesentlichen saisonalen Anteil besitzt, d.h. teste

$$H_0 : S_t \equiv 0 \quad \forall t \qquad \text{vs.} \qquad H_1 : S_t \not\equiv 0 \text{ für ein t.}$$

Dafür nehmen wir an, dass X_t, $t \in \mathbb{Z}$ uiv. $N(0, \sigma_x^2)$ Zufallsvariablen sind, $\sigma_x^2 > 0$.
Sei $x_j = \frac{2\pi j}{n},\ j \in \{1, \dots, m\},\ m = [\frac{n-1}{2}]$. Dann lautet die Teststatistik

$$T = \frac{\max\limits_{j=1,\dots,m} I_n(x_j)}{\sum_{j=1}^{m} I_n(x_j)} \qquad \in [0; 1].$$

Satz 3.4.1 (Satz von Fisher)
Unter den obigen Voraussetzungen gilt $T \overset{H_0}{\sim} F$, wobei F die Dichte

$$m(m-1) \sum_{j=1}^{\left[\frac{1}{x}\right]m} (-1)^{j-1} \binom{m-1}{j-1} (1-jx)^{m-z}$$

für $x \in [0, 1]$ besitzt.
Ohne Beweis.

Falls F bekannt ist, können Quantile F_γ^{-1} von F berechnet werden.
Entscheidungsregel: H_0 wird abgelehnt, falls $T > F_{1-\gamma}^{-1}$, wobei $\gamma \in (0, 1/2)$ klein ist (Wahrscheinlichkeit des Fehlers 1. Art).

Im Folgenden sollen (nach wie vor unter der Annahme bekannter $Z_1, \dots, Z_n$) aus dem Periodogramm Werte für b_k und φ_k mit Hilfe des folgenden Regressionsansatzes geschätzt

werden (dabei werden u und $\mu_k \quad k = 1, \ldots, n$ als bekannt vorausgesetzt). Benutzt wird dabei, dass $\cos(\alpha + \beta) = \cos\alpha\cos\beta - \sin\alpha\sin\beta$.

$$\begin{aligned} \Rightarrow Z_t &= \sum_{k=1}^{u} b_k \cos(\mu_k t)\cos(\varphi_k) - \sum_{k=1}^{u} b_k \sin(\mu_k t)\sin(\varphi_k) + X_t \ , \quad t = 1, \ldots, n \\ &= \sum_{k=1}^{u} \cos(\mu_k t)\underbrace{b_k \cos(\varphi_k)}_{\beta_k} - \sum_{k=1}^{u} \sin(\mu_k t)\underbrace{b_k \sin(\varphi_k)}_{\beta_k} + X_t \ , \quad t = 1, \ldots, n \end{aligned}$$

$$\underbrace{\begin{pmatrix} Z_1 \\ \vdots \\ Z_n \end{pmatrix}}_{\widetilde{Z}} = \underbrace{\begin{pmatrix} \cos(\mu_1) & \cdots & \cos(\mu_n) & -\sin(\mu_1) & \cdots & -\sin(\mu_u) \\ \cos(2\mu_1) & \cdots & \cos(2\mu_n) & -\sin(2\mu_1) & \cdots & -\sin(2\mu_u) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \cos(n\mu_n) & \cdots & \cdots & \cdots & \cdots & -\sin(n\mu_u) \end{pmatrix}}_{\text{Matrix } A} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_u \end{pmatrix} + \begin{pmatrix} X_1 \\ \vdots \\ X_u \end{pmatrix}$$

Daraus werden mit Hilfe der MKQ-Methode Schätzer $\hat{\beta}_k$ für $\beta = (\ldots)$ gewonnen.

$$\hat{\beta} = (\widehat{b_1 \cos\varphi_1}, \ldots, \widehat{b_n \cos\varphi_n}, -\widehat{b_1 \sin\varphi_1}, \ldots, -\widehat{b_n \sin\varphi_n})$$

Es ist also: $\hat{\beta} = (A^T A)^{-1} A^T \widetilde{Z}$, falls $rang(A_n) = 2n$. Es kann gezeigt werden, dass $\exists n_0 \in \mathbb{N} : \forall n \geq n_0 : \det(A_n) \neq 0$.

Lemma 3.4.1

Der Schätzer $\hat{\beta}$ für β ist erwartungstreu und konsistent, falls der Prozess $X = \{X_t,\ t \in \mathbb{Z}\}$ den oben beschriebenen Voraussetzungen genügt. Dann ist $\tilde{\beta}$ konsistent. Dabei können b_k und φ_k wie folgt geschätzt werden:

$$\begin{aligned} \widehat{\varphi_k} &= \arctan\left(\frac{\widehat{b_k \cos\varphi_k}}{\widehat{b_k \sin\varphi_k}}\right) \ , \quad k = 1, \ldots, n, \\ \widehat{b_k} &= \frac{\widehat{b_k \cos\varphi_k}}{\widehat{\cos\varphi_k}} = \frac{\widehat{b_k \sin\varphi_k}}{\widehat{\sin\varphi_k}} \ , \quad k = 1, \ldots, n, \end{aligned}$$

vorausgesetzt, dass sowohl $\widehat{\cos\varphi_k} \neq 0$, $\widehat{\sin\varphi_k} \neq 0$ als auch $\widehat{\cos\varphi_k} \neq 0$.

3.4.1 Zeitreihen mit polynomialem und saisonalem Anteil

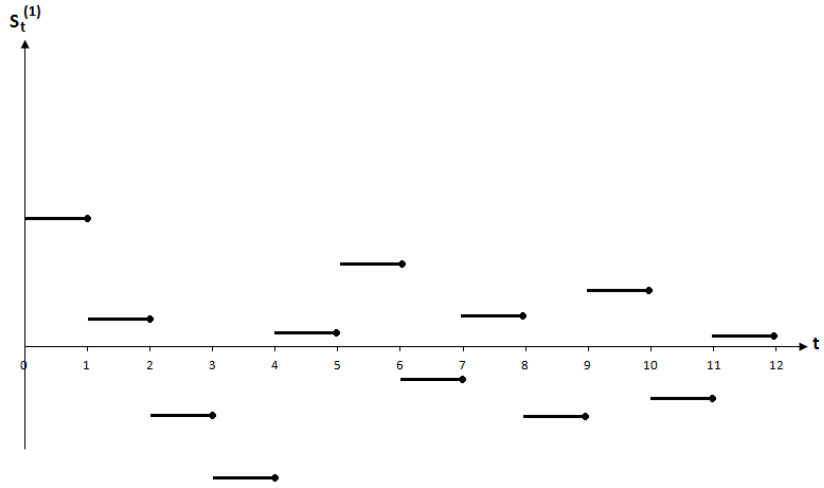
Sei die Zeitreihe $Z = \{Z_t : t \in \mathbb{Z}\}$ gegeben: $Z_t = T_t + S_t + X_t,\ t \in \mathbb{Z}$, wobei $T_t = \sum_{j=0}^{r} e_j t^j$ der polynomiale Anteil und $S_t = S_t^{(1)} + \cdots + S_t^{(n)}$ der saisonale Anteil ist mit einer periodischen Funktion $S^{(i)} : \mathbb{R} \to \mathbb{R}$ mit Periode $p^{(i)}$ und $\int_0^{p^{(i)}} S_x^{(i)} dx = 0$.

Bemerkung 3.4.1

Falls $\int_0^{p^{(i)}} S_x^{(i)} dx \neq 0$, so kann sein Wert zu c_0 im polynomiellen Anteil gerechnet werden.
Sehr oft: $S_t^{(1)} = \sum_{k=1}^{p^{(1)}} d_k I_{(e^{p^{(i)}}+k=1)}$

Beispiel 3.4.1

Im Beispiel 3.1.1 war $p^{(1)} = 12$ (Monate).



Dabei ist

$$S_t^2 + \cdots + S_t^{(n)} \equiv 0 \quad \forall t \quad \text{oder}$$
$$= \sum_{t=2}^{n} b_k \cos(\varphi_k + \mu_k t)$$

$X = \{X_t,\ t \in \mathbb{Z}\}$ ist ein ARMA(p,q)-Prozess, der die obigen Bedingungen erfüllt.

Zielsetzung: Schätze $e_0, \ldots, e_r, d_1, \ldots, d_{p^{(1)}}, b_k, \varphi_k, \mu_k,\ k = 1, \ldots, n$ aus den Beobachtungen $Z_1, \ldots, Z_n$.

1. Schätze $e_0, \ldots, e_r$ aus dem Regressionsmodell

$$Z_t = \sum_{j=0}^{r} e_j t^j + \underbrace{S_t + X_t}_{\varepsilon_t}, \quad t = 1, \ldots, n, \quad n \geq r+1,$$

wobei $r \in \mathbb{N}$ fest ist. Das heißt

$$\widetilde{Z} = B_n \cdot e + \varepsilon \ , \quad \text{wobei } \widetilde{Z} = (Z_1, \ldots, Z_n)^T, e = (e_0, \ldots, e_r)^T,$$
$$\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)^T$$
$$B_n = \begin{pmatrix} 1 & 1 & \ldots & 1 \\ 1 & 2 & \ldots & 2^r \\ \vdots & \vdots & \ddots & \vdots \\ 1 & n & \cdots & n^r \end{pmatrix} \ , \quad rg(B_n) = r+1$$

(Für $n = r+1$ ist B_{r+1} die Vandermonde-Matrix, die nicht ausgeartet ist.)

$\Longrightarrow \tilde{e} = (\tilde{e_0}, \ldots, \tilde{e_r})^T = (B_n^T B_n)^{-1} N_n^T \widetilde{Z}$ ist der MKQ-Schätzer für die Koeffizienten e. Dann gilt ferner $\sum_{j=0}^r \widetilde{e_j} t^j \approx T_t = \sum_{j=0}^r e_j t^j, \ t \in \mathbb{Z}$.

2. Bilde die Zeitreihe $\bar{Z} = \{\bar{Z}_t, \ t \in \mathbb{Z}\}$ durch $\bar{Z}_t = Z_t - \sum_{j=0}^r e_j t^j$ und unterstelle das Modell

$$\bar{Z}_t \approx S - t + X_t, \ t \in \mathbb{Z},$$

wobei $S_t = \sum_{k=1}^u b_k \cos(\varphi_k + \mu_k t), \ t \in \mathbb{Z}$.

Schätze daraus u und μ_k, $k = 1, \ldots, u$ wie in §3.4.1 beschrieben (benutze dabei das Periodogramm von $\bar{Z}$). $\tilde{e}$ wird im weiteren Verlauf nicht mehr verwendet.

3. Verwende das lineare Regressionsmodell

$$Z_t = \sum_{j=0}^r e_j t^j + \sum_{k=1}^u \cos(\mu_k t) b_k \cos(\varphi_k) - \sum_{k=1}^u \sin(\mu_k t) b_k sin\varphi_k + X_t \ , \quad t = 1, \ldots, n.$$

In Matrixschreibweise lautet dieses

$$\begin{aligned} \widetilde{Z} &= C_n \cdot \beta + \varepsilon \ , \quad \text{wobei} \\ \widetilde{Z} &= (Z_1, \ldots, Z_n)^\top \\ \varepsilon &= (X_1, \ldots, X_n)^\top \\ \beta &= (e_0, \ldots, e_r, b_1 \cos\varphi_1, \ldots, b_n \cos\varphi_n, -b_1 \sin\varphi_1, \ldots, b_n \sin\varphi_n)^\top \\ C_n &= (B_n A_n) \end{aligned}$$

Für alle $n \geq n_0$ mit einem $n_0 \in \mathbb{N}$ ist $rg(C_n) = 2n + r + 1$.
$\Rightarrow$ Der MKQ-Schätzer lautet

$$\widehat{\beta} = (\widehat{e_0}, \ldots, \widehat{e_r}, b_1 \widehat{\cos\varphi_1}, \ldots, b_u \widehat{\cos\varphi_u}, -\widehat{b_1 \cos\varphi_1}, \ldots, b_u \widehat{\cos\varphi_u})$$

Genau wie im Lemma 3.4.1 kann gezeigt werden, dass $\hat{\beta}$ ein erwartungstreuer und konsistenter Schätzer für β ist. Dabei gilt $\widehat{\varphi_k} = \arctan \frac{\widehat{b_k \sin\varphi_k}}{\widehat{b_k \cos\varphi_k}}$, $\quad \widehat{b_k} = \frac{\widehat{b_k \cos\varphi_k}}{\widehat{\cos\varphi_k}} = \frac{\widehat{b_k \sin\varphi_k}}{\widehat{\sin\varphi_k}}$,
$k = 1, \ldots, n$ falls der Schätzer $b_k \widehat{\cos\varphi_k} \neq 0$ und $(\widehat{\sin\varphi_k}) \neq 0$. Dafür ist es ausreichend, vorauszusetzen, dass $b_k \neq 0$, $\varphi_k \neq 0$ und n groß genug (wegen Konsistenz von $\hat{\beta}$).

Satz 3.4.2

Ähnlich wie in §3.4.1 und 3.4.2 können Parameter $d_1, \ldots, d_{p^{(1)}}$ von Zeitreihen mit dem Treppenfunktion-Anteil $S^{(1)}$ mit Hilfe von MKQ-Schätzern im entsprechenden linearen multivariaten Regressionsmodell geschätzt werden:

$$Z_t = T_t + S_t + X_t \ , \quad t = 1, \ldots, n \ .$$

4 Appendix

4.1 Regression

4.1.1 Einfache lineare Regression

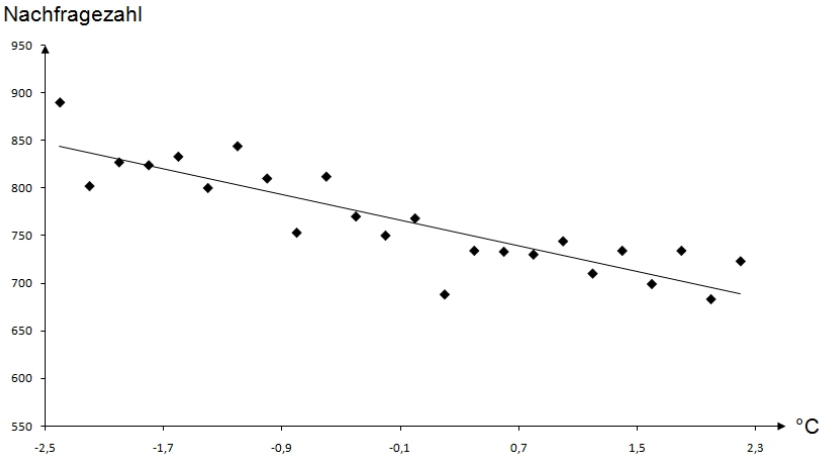
Betrachten wir das folgende lineare Modell

$$Y_i = \alpha + x_i\beta_i + \varepsilon_i \ , \quad i = 1, \ldots, n \ ,$$

wobei Y_i und $\varepsilon_i,\ i = 1, \ldots, n$ Zufallsvariablen sind. $x_1, \ldots, x_n$ 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 $Y_1, \ldots, Y_n$ und $x_1, \ldots, x_n$.

Bemerkung: Die Störgrößen $\varepsilon_i, \ldots, \varepsilon_n$ 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 $y = \alpha + x\beta$:

$$e(\alpha, \beta) = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \alpha - \beta x_i)^2 \to \min_{\alpha,\beta}$$

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:

$$\hat{\beta} = \frac{s^2_{xy}}{s^2_{xx}} \ , \quad \hat{\alpha} = \overline{Y}_n - \hat{\beta}\overline{x}_n \ , \tag{4.1.5}$$

wobei

$$s^2_{xx} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x}_n)^2 \quad \text{die Strichprobenvarianz von } (x_1,\ldots,x_n) \ ,$$
$$s^2_{yy} = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \overline{y}_n)^2 \quad \text{die Strichprobenvarianz von } (Y_1,\ldots,Y_n) \ ,$$
$$s^2_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x}_n)(y_i - \overline{y}_n) \quad \text{die Strichprobenkovarianz von } (x_1,\ldots,x_n) \text{ und } (Y_1,\ldots,Y_n)$$

und

$$\overline{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \text{das arithmetische Mittel von } (x_1,\ldots,x_n) \ ,$$
$$\overline{y}_n = \frac{1}{n}\sum_{i=1}^{n} y_i \quad \text{das arithmetische Mittel von } (Y_1,\ldots,Y_n)$$

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:

$$Y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_m x_{im} + \varepsilon_i \ , \quad i = 1, \ldots, n \ , \quad m \ll n \ ,$$

oder in Matrix-Form

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = \underbrace{\begin{pmatrix} x_{11} & \ldots & x_{1m} \\ \vdots & & \vdots \\ x_{n1} & \ldots & x_{nm} \end{pmatrix}}_{X} \underbrace{\begin{pmatrix} \beta_1 \\ \vdots \\ \beta_m \end{pmatrix}}_{\beta} + \underbrace{\begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}}_{\varepsilon}$$

$$\text{bzw.} \quad Y = X\beta + \varepsilon \quad \text{in Kurzschreibweise ,}$$

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)
- $\varepsilon_i, \ldots, \varepsilon_n$ unkorreliert: $\mathrm{Cov}\,(\varepsilon_i, \varepsilon_j) = \delta_{ij}\sigma^2 = \begin{cases} \sigma^2 & , \quad i = j \ , \\ 0 & , \quad i \neq j \ . \end{cases}$
- $\mathrm{Var}\,\varepsilon_i = \mathbb{E}\,\varepsilon_i^2 - (\mathbb{E}\,\varepsilon_i)^2 = \mathbb{E}\,\varepsilon_i^2 = \sigma^2 \ , \quad i = 1, \ldots, n$

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:

$$e(\beta) = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \beta_1 x_{i1} - \ldots - \beta_m x_{im})^2 \to \min_{\beta}$$

Ähnlich wie in Abschnitt 4.1.1 ist der MKQ-Schätzer $\hat{\beta}$ eine Lösung der sog. *Normalengleichung*

$$(X^\top X)\beta = X^\top Y \ .$$

Falls $rg(X) = m$, dann (und genau dann) ist $X^\top X$ eine invertierbare Matrix und es gilt

$$\hat{\beta} = (X^\top X)^{-1} X^\top Y \ .$$

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:

$$\begin{aligned}\mathbb{E}\,(\hat{\beta}) &= (X^\top X)^{-1} X^\top \mathbb{E}\, Y = (X^\top X)^{-1} X^\top \mathbb{E}\,(X\beta + \varepsilon) \\ &= \underbrace{(X^\top X)^{-1}(X^\top X)}_{\mathrm{Id}}\beta + (X^\top X)^{-1} X^\top \underbrace{\mathbb{E}\,\varepsilon}_{=0} = \beta\end{aligned}$$

□

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:

$$Y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} \ , \quad i = 1, \ldots, 9$$

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.

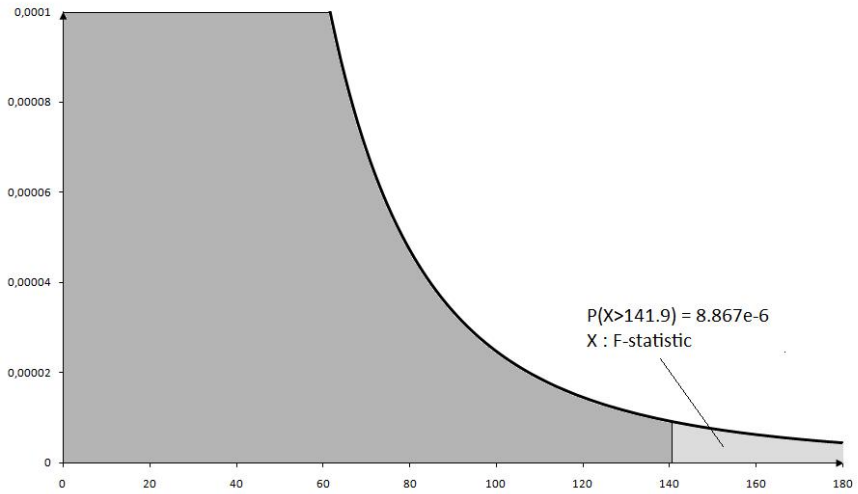
```
Call:
lm(formula = Besucherzahl ~ 1 + Pistenlaenge + Liftkapazitaet)

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
```

Da der p-Wert = 8.867e-6 ist, ist dies ein sehr stark signifikanter Hinweis darauf, dass



Dichte der $F_{2,6}$ - Verteilung

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

AIC-Kriterium, 54
asymptotische Tests, 46
Autokorrelationsfunktion, 73
Autokovarianzfunktion, 73

Bestimmtheitsmaß R^2, 30
Bochner
    Satz von, 77
Box-Cox-Test, 33
Box-Cox-Transformation, 21
Box-Cox-Verfahren, 33

exogene Variablen, 6
Exponentialfamilie, 37

Fisher Scoring, 46, 51
Fisher-Informationsmatrix, 42, 52

Gauß-Newton-Verfahren, 13

Hesse-Matrix, 43
Hilbertraum, 66

Informationskoeffizient von Akaike, 54

Korrelationsfunktion, 73
Kovarianzfunktion, 73

Likelihood-Ratio-Test, 34, 35, 53
Lineare Approximation, 10
Linkfunktion, 37
    natürliche, 39
Log-Likelihood-Gleichungen
    konzentrierte, 20
Logit-Modell, 40, 51, 56

Maximum-Likelihood-Methode, 18, 24
Methode der kleinsten Quadrate, 99
mittlerer quadratischer Abstand, 99
Modelle
    quasilineare, 8
    verallgemeinerte lineare, 37

Newton-Verfahren, 45
Normalengleichung, 101

Odd, 40

p-Wert, 102
Paneldaten, 6
Poisson-Modell, 56
Poisson-Regression, 44
Probit-Modell, 40
Prozess
    autoregressiver, AR(1), 59
    der gleitenden Mittel, 68
    linearer, 66

Quantilfunktion der Normalverteilung, 40
Querschnittsdaten, 6

Regression
    binäre kategoriale, 40
    einfache lineare, 8, 99
    logistische, 40, 44
    multivariate lineare, 101
Regressoren, 6

saisonaler Anteil, 62
Score-Funktion, 52
Score-Statistik, 53
Spektraldichte, 78
spektraler Ansatz, 80
Spektralfunktion, 78

Trend, 62

Wald-Statistik, 53
Wold
    Zerlegungssatz von, 70

Yule-Walker-Gleichung, 76

Zeitreihen, 5, 58, 62