

Dr. Benedikt Weber

Leugrueb 10

8126 Zumikon

**Vorhersage des Verhaltens von Talsperren
mit Hilfe des Soll-Ist-Vergleichs —
Statistischer Teil**

Prédiction du comportement des barrages par
comparaison mesuré-calculé — Aspects statistiques

Im Auftrag des Bundesamtes für Wasser und Geologie, Biel

Zumikon, den 4. April 2002

Inhaltsverzeichnis

Inhaltsverzeichnis	i
Vorwort	v
1 Einleitung	1
1.1 Motivation und Zielsetzung	1
1.2 Übersicht über statistische Methoden.....	1
1.3 Abgrenzung.....	3
1.4 Berechnungen	3
2 Einfache lineare Regression.....	4
2.1 Methode der kleinsten Quadrate	4
2.1.1 Problemstellung	4
2.1.2 Berechnung der Koeffizienten	4
2.2 Statistische Betrachtung	5
2.2.1 Koeffizienten als Zufallsvariablen	5
2.2.2 Varianzanalyse (Analysis of Variance, ANOVA).....	8
2.2.3 Korrelationskoeffizient	9
2.2.4 F-Test	10
2.3 Signifikanz und Konfidenzintervalle.....	10
2.3.1 t-Verteilung	10
2.3.2 t-Test	11
2.3.3 Zusammenhang mit F-Test	12
2.3.4 Konfidenzintervall für Regressionskoeffizienten.....	13
2.3.5 Konfidenzintervall für Erwartungswert von y	13
2.3.6 Vorhersageintervall (Prediction interval)	13
2.4 Überprüfung der Annahmen	14
2.4.1 Normalverteilung der Residuen.....	14
2.4.2 Heteroskedastizität und Autokorrelation	15
3 Multiple Regression	18
3.1 Methode der kleinsten Quadrate	18
3.2 Varianzanalyse	19
3.2.1 Quadratsummen.....	19
3.2.2 Korrelationskoeffizient	20
3.2.3 Globaler F-Test.....	20
3.3 Partieller F-Test, partieller Korrelationskoeffizient	21
3.3.1 Partieller F-Test bei Elimination	21
3.3.2 Partieller F-Test bei Einfügen, partieller Korrelationskoeffizient.....	22
3.4 Signifikanz und Konfidenzintervalle.....	23
3.4.1 t-Test	23
3.4.2 Konfidenz- und Prognoseintervalle	23
3.4.3 Extrapolation	24
3.5 Multikollinearität.....	25
3.5.1 Variable Inflation Factor, VIF.....	25

3.5.2	Auswirkungen der Multikollinearität	26
3.6	Variablenauswahl	26
3.6.1	Elimination	26
3.6.2	Einfügen	27
4	Weitergehende Verfahren.....	29
4.1	Standardisierte Variablen.....	29
4.2	Principal-Component-Regression	30
4.2.1	Methode	31
4.2.2	Beispiel	32
4.2.3	Interpretation	33
4.2.4	t-Test für transformierte Regressoren.....	34
4.2.5	Ursprüngliche Regressionskoeffizienten und t-Test.....	35
4.2.6	Diskussion	36
4.3	Ridge-Regression	37
4.3.1	Diskussion	38
4.4	Autokorrelation der Fehler	38
4.4.1	Diskussion	41
5	Regressionsmodelle für Staumauern.....	43
5.1	Einflussgrößen	43
5.2	Polynome	43
5.3	Periodische Einflüsse (Saisonale Einflüsse)	44
5.4	Wärmeleitung.....	45
5.5	Kriechen.....	48
5.6	Irreversible Verformungen (Drift)	48
6	Anwendung auf Staumauer Schlegeis	50
6.1	Beschreibung.....	50
6.1.1	Überblick.....	50
6.1.2	Messdaten.....	51
6.2	Anmerkung zu Tabellen und Plots	53
6.3	Klassisches Modell ohne Temperaturbeobachtungen	53
6.3.1	Methode der kleinsten Quadrate.....	54
6.3.2	Autokorrelation der Fehler.....	56
6.3.3	Abhängigkeit von der Regressionsperiode	59
6.3.4	Polynome und Skalierung	60
6.4	Modell mit Betontemperaturen	62
6.4.1	Beseitigung der Multikollinearität durch Regressor-Elimination	63
6.4.2	Beseitigung der Multikollinearität durch Principal-Component-Regression	64
6.4.3	Autokorrelation der Fehler.....	67
6.5	Modell mit Wärmeleitung.....	69
6.5.1	Klassische Regression und Ridge-Regression	69
6.5.2	Principal-Component-Regression.....	71
6.5.3	Autokorrelation der Fehler	72
6.6	Weitere Modelle.....	74
6.6.1	Model mit Kriechen.....	74
6.6.2	Modell mit Lufttemperatur.....	74
6.7	Zusammenfassung Schlegeis	75
7	Anwendung auf Staumauer Sambuco	77
7.1	Beschreibung.....	77

7.1.1	Übersicht	77
7.1.2	Beschreibung der Messdaten.....	78
7.2	Klassisches Modell ohne Temperaturen.....	78
7.2.1	Methode der kleinsten Quadrate.....	79
7.2.2	Autokorrelation der Fehler	81
7.2.3	Abhängigkeit von der Regressionsperiode	83
7.3	Modell mit Betontemperaturen	83
7.3.1	Methode der kleinsten Quadrate.....	83
7.3.2	Autokorrelation der Fehler	84
7.4	Modell mit Auftriebsdrücken.....	85
7.4.1	Methode der kleinsten Quadrate.....	85
7.4.2	Autokorrelation der Fehler	87
7.5	Modell mit Wärmeleitung.....	88
7.5.1	Methode der kleinsten Quadrate.....	88
7.5.2	Autokorrelation der Fehler	89
7.6	Zusammenfassung Sambuco.....	91
7.7	Anhang: Plots der Temperaturen und Auftriebsdrücke	93
8	Beispiel mit Langzeitverformungen	96
8.1	Beschreibung.....	96
8.2	Langzeitprognose	97
8.2.1	Klassische Methode	97
8.2.2	Modell mit hohen VIF-Werten.....	98
8.3	Regression über ganze Messperiode	99
8.3.1	Modell ohne Langzeitfunktionen	99
8.3.2	Modell mit Exponentialfunktionen.....	99
8.4	Autokorrelation der Fehler	101
8.5	Zusammenfassung Beispiel mit Langzeitverformungen.....	102
9	Zusammenfassung und Schlussfolgerungen.....	103
9.1	Modellierung und Einflussfunktionen.....	103
9.1.1	Stauhöhe.....	103
9.1.2	Periodische (saisonale) Effekte	103
9.1.3	Temperatur	103
9.1.4	Kriechen	104
9.1.5	Irreversible Effekte (Drift)	104
9.2	Statistische Grössen.....	105
9.2.1	Beurteilungsgrössen für das gesamte Modell	105
9.2.2	Beurteilungsgrössen für einzelne Regressoren	105
9.2.3	Multikollinearität	106
9.2.4	Prognoseintervall und Extrapolation	106
9.2.5	Überprüfung der Annahmen	107
9.3	Variablenauswahl	107
9.4	Weitergehende Methoden.....	108
9.4.1	Principal-Component-Regression.....	108
9.4.2	Ridge-Regression	108
9.4.3	Verallgemeinerte Methode der kleinsten Quadrate	109
9.5	Allgemeine Folgerungen	110
A	Mathematische Herleitungen.....	111
A.1	Inverse von Blockmatrizen	111
A.2	SS_{Res}	111

A.3	VIF	111
A.4	Regressor eliminieren oder einfügen, partieller F-Test	112
A.4.1	Partieller F-Test bei der Elimination	112
A.4.2	Partieller F-Test beim Einfügen	113
A.5	Standardisierte Variablen.....	115
A.5.1	Zentrierung	115
A.5.2	Skalierung	117
A.6	Partieller F-Test für Principal-Component-Regression.....	118
A.7	Numerische Behandlung der Methode der kleinsten Quadrate.....	119
A.7.1	Formulierung mit QR-Zerlegung	120
A.7.2	Singulärwertzerlegung	120
A.7.3	Singulärwertzerlegung und Principal-Component-Regression	121
A.8	Verallgemeinerte Methode der kleinsten Quadrate.....	122
A.8.1	Allgemeine Theorie	122
A.8.2	Schätzung des Autokorrelationsparameters	124
Literatur	125

Vorwort

Der Soll-Ist-Vergleich ist ein wichtiges Mittel bei der Überwachung von Talsperren, das erlaubt ein unregelmässiges Verhalten frühzeitig festzustellen. Er wird auch in der Schweiz seit Jahren zu diesem Zweck angewandt. Eine häufig verwendete Methode ist die statistische Methode, bei der empirische Einflussfunktionen benützt und die entsprechenden Koeffizienten aufgrund früherer Beobachtungen ermittelt werden.

Die Arbeitsgruppe Numerik des Schweizerischen Talsperrenkomitees hat eine umfassende Übersicht über die verschiedenen Ansätze und Methoden des Soll-Ist-Vergleichs erarbeitet und ist dabei auch auf einige offene Fragen, insbesondere im Bezug auf die Statistik, gestossen. Um diese Lücke zu schliessen, hat das Bundesamt für Wasser und Geologie (BWG), Sektion Talsperren, diese Studie in Auftrag gegeben. Das Ziel der Studie war, die neueren Erkenntnisse der Statistik in den Soll-Ist-Vergleich einzubeziehen. Die Studie wurde von Seiten des BWG von den Herren Dr. G.R. Darbre und A. Kobelt geleitet und begleitet. Der Verfasser dankt den beiden Herren für die gute Zusammenarbeit und die wertvollen Anregungen und Diskussionen.

Die vorliegende Studie bestätigt einerseits die Richtigkeit der bisherigen Praxis, gibt aber auch einige wichtige Erweiterungen, die die Beurteilung eines Modells und eine rationale Auswahl der Einflussfunktionen erlauben. Der Verfasser hofft mit dieser theoretischen Studie auch für die Praxis einen wichtigen Beitrag zu liefern.

Zumikon, im April 2002

B. Weber

1 Einleitung

1.1 Motivation und Zielsetzung

Ein wichtiges Mittel zur Überwachung von Talsperren ist die laufende Erfassung von Verschiebungen und anderen Messgrößen und deren Vergleich mit zu erwartenden Werten. Damit kann eine Abweichung vom normalen Verhalten frühzeitig festgestellt werden. Die Vorhersage erfolgt anhand von Modellen, die aus früher erhobenen Messdaten und aus numerischen Berechnungen bestimmt werden. Die Arbeitsgruppe “Numerik” des Schweizerischen Talsperrenkomitees hat in einem Bericht die wichtigsten Methoden und Ansätze beschrieben (Schweizerisches Talsperrenkomitee 2000). Alle Modelle, ausser den rein deterministischen, beruhen auf einer linearen Regression mit vorgewählten Einflussfunktionen. Die Wahl dieser Einflussfunktionen bleibt der Erfahrung und der Intuition des Ingenieurs überlassen. Hier kann die Statistik einen wesentlichen Beitrag leisten. Mit ihrer Hilfe kann der Ingenieur beurteilen, welche Einflussfunktionen wichtig sind und welche nicht. So können Modelle aufgestellt werden, die eine bessere Prognose erlauben und mit denen sich eine Abweichung vom normalen Verhalten früher feststellen lässt.

In dieser Studie soll die lineare Regression mit ihren statistischen Methoden auf die Überwachung von Talsperren angewandt werden. Es soll festgestellt werden, welche Standard-Methoden für diese Aufgabe geeignet sind, und wie sie konkret angewendet werden können. Das beinhaltet sowohl eine Beschreibung der relevanten Theorie, wie auch das Berechnen von entsprechenden Beispielen. Die ursprüngliche Idee, ein rezeptartiges Vorgehen für den praktischen Anwender zu geben, konnte nur für die ganz einfachen Fälle realisiert werden. Für die komplizierteren Fälle gibt es meist mehrere Strategien, die von Fall zu Fall verschieden gut funktionieren. Hier ist noch viel mehr Erfahrung mit verschiedenen Talsperren nötig, um allgemeine Empfehlungen abgeben zu können.

Bei den Untersuchungen ist es wichtig, sich den Zweck der Berechnungen immer vor Augen zu halten. Es geht nicht um eine möglichst genaue Regression, sondern um eine Prognose. Aber es geht auch nicht um eine Prognose in dem Sinn, dass vorausgesagt werden soll, wie sich eine Talsperre verhalten wird, sondern wie sie sich unter normalen Umständen verhalten sollte. Durch den Vergleich mit dem tatsächlichen Verhalten können dann Abweichungen frühzeitig festgestellt werden.

1.2 Übersicht über statistische Methoden

Die klassische Methode zur Bestimmung der Regressionskoeffizienten ist die Methode der kleinsten Quadrate. Die Regressionskoeffizienten werden so bestimmt, dass die Quadratsumme der Residuen minimal wird. In der statistischen Betrachtungsweise

werden die Regressionskoeffizienten aber nicht als feste Grössen, sondern als Zufallsvariablen angesehen. Der berechnete Wert ist nur eine Schätzung des Mittelwertes. Aus der statistischen Verteilung der Eingabedaten kann die Varianz der Regressionskoeffizienten berechnet werden, und mit der t -Testgrösse (mit einer Student-Verteilung) kann bestimmt werden, ob die einzelnen Variablen statistisch signifikant sind. Die t -Kenngrosse ist auch ein Mass dafür, wie stark die Quadratsumme der Residuen zunimmt, wenn die entsprechende Variable aus dem Modell eliminiert wird. Damit ist die t -Kenngrosse die wichtigste Grösse zur Bewertung der einzelnen Variablen in einem Modell.

Eine weitere wichtige Grösse ist der VIF (Variance Inflation Factor). Er deckt auf, wenn eine so genannte Multikollinearität existiert, d.h. wenn zwei oder mehrere Variablen stark voneinander abhängig sind. Die Regressionskoeffizienten können dann nicht mehr individuell sondern nur noch als Kombination bestimmt werden. Die Multikollinearität verfälscht auch die t -Kenngrossen, was die Variablenauswahl erschwert.

Eine effiziente Methode zur Beseitigung der Multikollinearität ist die Principal-Component-Regression. Es handelt sich um eine Eigenwertzerlegung der Korrelationsmatrix. Durch Weglassen von Eigenvektoren mit kleinen Eigenwerten wird die Multikollinearität beseitigt. Für die Variablenauswahl mit einer Multikollinearität ist aber oft die Ridge-Regression besser geeignet. Die Ridge-Regression ist eine einfache Modifikation der Korrelationsmatrix, sodass diese weniger stark singulär wird.

Ein weiterer wichtiger Aspekt ist die Überprüfung der Annahmen der linearen Regression. Die wichtigsten Annahmen sind, dass die Fehler eine konstante Varianz haben und dass sie voneinander unabhängig sind. Die häufigste Verletzung der letzteren Annahme ist bei Zeitreihen die Autokorrelation der Fehler. Eine solche kann mit dem Korrelogramm oder anhand des Durbin-Watson-Tests festgestellt werden. Eine Autokorrelation der Fehler hat zur Folge, dass zwar die Regressionskoeffizienten mit der Methode der kleinsten Quadrate immer noch konsistent berechnet werden, dass aber die Varianz der Regressionskoeffizienten unterschätzt und damit die t -Kenngrossen überschätzt wird. Praktisch heisst das, dass man die Methode der kleinsten Quadrate immer noch anwenden kann, dass aber gewisse Variablen als statistisch signifikant ausgewiesen werden, obwohl sie es gar nicht sind. Die t -Kenngrossen sind also nicht mehr ein absolutes Auswahlkriterium, sondern nur noch ein relatives, indem sie angeben welche Variablen die Residuen am wenigsten beeinflussen.

Obwohl die Verletzung der Annahme der Unabhängigkeit der Fehler keine gravierenden Konsequenzen hat, ist es trotzdem wünschenswert, die Theorie so zu erweitern, dass sie eine Autokorrelation der Fehler berücksichtigt. Dies führt zur verallgemeinerten Methode der kleinsten Quadrate. Im Fall, dass die Fehler durch ein autoregressives Modell erster Ordnung beschrieben werden können, reduziert sich diese Methode darauf, dass man die abhängige und die unabhängigen Variablen mit einem verallgemeinerten Differenzenschema transformiert. Die transformierten Grössen erfüllen dann die Annahmen der klassischen Methode der kleinsten Quadrate, sodass diese angewandt werden kann. Beispiele zeigen, dass mit dieser Methode die Regressionskoeffizienten nur wenig ändern und damit die Quadratsumme der Residuen nur leicht zunimmt, dass

aber die t -Kenngrößen viel aussagekräftiger sind für die Modellbildung. Zusätzlich verringert sich auch die Multikollinearität und die Prognosen werden in manchen Fällen besser. Als Nachteil muss erwähnt werden, dass manchmal numerische Probleme auftreten können.

1.3 Abgrenzung

Die Untersuchungen beschränken sich auf die lineare Regression, d.h. auf Ansätze bei denen die Regressionskoeffizienten als lineare Unbekannte vorkommen. Wie üblich wurde dabei angenommen, dass die unabhängigen Variablen gegeben sind, und nur die abhängige Variable y eine Zufallsvariable ist. Fehler in den unabhängigen Variablen wurden nicht berücksichtigt.

Die klassische Methode der kleinsten Quadrate setzt voraus, dass die Fehler unkorreliert sind. Da diese Annahme in der Anwendung häufig nicht erfüllt ist, wurde zusätzlich die Autokorrelation der Fehler durch die verallgemeinerte Methode der kleinsten Quadrate berücksichtigt. Hier wurde aber nur der einfachste Fall untersucht, nämlich die Autokorrelation erster Ordnung.

Nicht untersucht wurden statistische Modelle mit autoregressiven Ansätzen, d.h. solche, bei denen frühere Werte der abhängigen Variablen als unabhängige Variable benutzt werden. Viele Grundsätze der klassischen linearen Regression gelten für solche Modelle nicht mehr, da in diesem Falle auch unabhängige Variablen als Zufallsvariablen betrachtet werden müssen.

Die Aufgabe dieser Studie war nicht die Entwicklung von neuen Einflussfunktionen, sondern die Anwendung mit Hilfe der Statistik. Trotzdem wurden einige neuartige Funktionen eingeführt, wie z.B. die Tschebyscheff-Polynome oder die Wärmeleitungsfunktionen. Es ist aber klar, dass hier noch ein weites Feld offen ist.

Die Beispiele beschränken sich auf drei Mauern, wobei jeweils nur eine Messgröße als unabhängige Variable gewählt wurde. Hingegen wurden zahlreiche Modelle mit verschiedenen unabhängigen Variablen berechnet. Trotzdem ist Anzahl der Beispiele zu klein, um in allen Fällen allgemeine Empfehlungen abgeben zu können. Dies trifft insbesondere für die weniger bekannten Methoden zu, wie die Principal-Component-Regression oder die verallgemeinerte Methode der kleinsten Quadrate.

1.4 Berechnungen

Um die Berechnungen durchzuführen, wurde ein Prototyp eines Computerprogramms geschrieben. Obwohl es zahlreiche gute Statistikprogramme gibt, wurde dieser Weg als sinnvoll erachtet. Einerseits muss man sich die einzelnen Schritte viel genauer überlegen, wenn man sie programmiert und andererseits kann man mit dem eigenen Programm viel freier herumexperimentieren. Wichtig für die Berechnungen war auch, dass mit wenig Aufwand zahlreiche Varianten ausprobiert werden konnten. Dazu wurden die Transformationen der Variablen, wie z.B. die einfache Wärmeleitung, direkt ins Programm eingebaut, so dass keine separate Vorberechnung nötig war.

2 Einfache lineare Regression

In diesem Kapitel wird eine kurze Einführung in die lineare Regression anhand der einfachen Regression gegeben, d.h. der Regression mit einer einzigen unabhängigen Variablen. Neben der üblichen Behandlung der Methode der kleinsten Quadrate wird vor allem auf die statistischen Aspekte eingegangen. Dieser Einführungstext folgt im Wesentlichen dem 2. Kapitel aus dem Buch Montgomery et al. (2001). Einige Aspekte wurden auch aus dem Buch von Pindyck und Rubinfeld (1998) übernommen. Das Hauptgewicht dieses Kapitels liegt auf Erklärungen und einer Zusammenfassung der wichtigsten Resultate. Für eine ausführlichere Herleitung wird der Leser auf die beiden Bücher verwiesen.

2.1 Methode der kleinsten Quadrate

2.1.1 Problemstellung

Eine Reihe von Beobachtungen (Messdaten) x_i und y_i soll durch eine lineare Funktion approximiert werden. Die Messdaten x_i werden dabei als unabhängige Grösse, die Daten y_i als abhängige Grösse betrachtet. Die unabhängige Grösse wird auch als Regressor bezeichnet. Die beobachteten Messwerte y_i liegen nicht genau auf einer Geraden, sondern sind mit einem Fehler ε_i behaftet. Somit kann die Beziehung geschrieben werden als

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i. \quad (2.1)$$

Die bekannteste Methode zur Bestimmung der Koeffizienten ist die Methode der kleinsten Quadrate.

2.1.2 Berechnung der Koeffizienten

Bei der Methode der kleinsten Quadrate werden die Regressionskoeffizienten so bestimmt, dass die Quadratsumme der Abweichungen minimiert wird. Um

$$S = \sum_{i=1}^n \varepsilon_i^2 = (y_i - \beta_0 - \beta_1 x_i)^2 \quad (2.2)$$

zu minimieren, werden die partiellen Ableitungen nach den Parametern β_0 und β_1 zu Null gesetzt. Daraus resultieren die geschätzten Koeffizienten $\hat{\beta}_0$ und $\hat{\beta}_1$ (Montgomery, S. 15)

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \quad \text{und} \quad \hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}. \quad (2.3)$$

Dabei sind \bar{x} und \bar{y} die Mittelwerte der Beobachtungswerte, d.h.

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad \text{und} \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i. \quad (2.4)$$

Die weiteren verwendeten Grössen sind definiert als

$$S_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2 \quad \text{und} \quad S_{xy} = \sum_{i=1}^n y_i (x_i - \bar{x}). \quad (2.5)$$

Aus Gleichung (2.3) ist ersichtlich, dass die Gerade $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ durch den Schwerpunkt (\bar{x}, \bar{y}) der Beobachtungswerte geht.

Die Differenz zwischen den beobachteten Werten y_i und den entsprechenden berechneten Werten \hat{y}_i sind die Residuen

$$e_i = y_i - \hat{y}_i \quad (2.6)$$

Eine wichtige Eigenschaft ist die, dass die Summe der Residuen verschwindet und damit ihr Mittelwert gleich Null ist:

$$\sum_{i=1}^n e_i = 0. \quad (2.7)$$

Ebenfalls kann gezeigt werden, dass die Residuen orthogonal zu den Werten \hat{y}_i sind:

$$\sum_{i=1}^n \hat{y}_i e_i = 0. \quad (2.8)$$

Geometrisch betrachtet ist deshalb die Methode der kleinsten Quadrate eine Projektion der beobachteten y_i auf den Raum der unabhängigen Variablen.

2.2 Statistische Betrachtung

2.2.1 Koeffizienten als Zufallsvariablen

In der bisherigen Betrachtung wurden die Beobachtungen als fixe gegebene Grössen betrachtet. Man kann sie aber auch als Stichprobe einer Grundgesamtheit betrachten. Die Gleichung

$$y = \beta_0 + \beta_1 x + \varepsilon \quad (2.9)$$

beschreibt dann die Grundgesamtheit, wobei y und ε jetzt Zufallsvariablen sind, die eine gewisse statistische Verteilung aufweisen. Es wird vorausgesetzt, dass der Fehler ε den Erwartungswert 0 und die noch unbekannte Varianz σ^2 hat. Über die Verteilung des Fehlers wird zunächst keine Annahme getroffen. Weiter wird vorausgesetzt, dass die Fehler unkorreliert sind.

In diesem Modell wird x nicht als Zufallsvariable angesehen, sondern als gegebene Grösse. Messfehler in der unabhängigen Variablen werden somit nicht berücksichtigt. Die Streuung der abhängigen Variablen y ist in dieser Betrachtung ebenfalls nicht in erster Linie auf Messfehler zurückzuführen, sondern beschreibt die Streuung des physikalischen Modells (z.B. Materialkennwerte) und anderer nicht berücksichtigter Einflüsse.

Der Erwartungswert (Mittelwert) von y bei gegebenem x ist damit

$$E(y | x) = \beta_0 + \beta_1 x \quad (2.10)$$

und die Varianz

$$\text{var}(y | x) = \text{var}(\beta_0 + \beta_1 x + \varepsilon) = \sigma^2. \quad (2.11)$$

Die Gerade $y = \beta_0 + \beta_1 x$ beschreibt also den Mittelwert der Zufallsvariablen y bei gegebenem x , während die statistische Verteilung durch ε beschrieben wird. Dieses statistische Modell ist in Bild 2-1 illustriert.

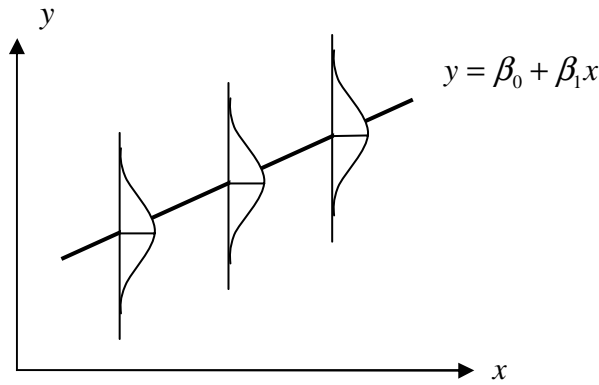


Bild 2-1: Regressionsgerade mit Verteilung der y-Werte

Aus der Methode der kleinsten Quadrate erhält man die geschätzten Regressionskoeffizienten $\hat{\beta}_0$ und $\hat{\beta}_1$. Da diese anhand einer Stichprobe berechnet werden, hängen sie von der Stichprobe ab und sind somit ebenfalls Zufallsvariablen mit einer gewissen Verteilung. Die Annahme, dass die Regressionskoeffizienten Zufallsvariablen sind, mag vorerst etwas ungewohnt erscheinen, ist aber wesentlich zum Verständnis. Es kann gezeigt werden (Montgomery, S.20), dass der Erwartungswert dieser Zufallsvariablen gerade den wahren Regressionskoeffizienten entspricht. Die Werte $\hat{\beta}_0$ und $\hat{\beta}_1$ sind unverzerrte Schätzwerte (unbiased estimate):

$$E(\hat{\beta}_0) = \beta_0 \quad \text{und} \quad E(\hat{\beta}_1) = \beta_1. \quad (2.12)$$

Ebenfalls kann die Varianz der Regressionskoeffizienten bestimmt werden:

$$\text{var}(\hat{\beta}_0) = \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right) \quad \text{und} \quad \text{var}(\hat{\beta}_1) = \frac{\sigma^2}{S_{xx}}. \quad (2.13)$$

Die Varianz σ^2 kann aus den Beobachtungsdaten geschätzt werden. Dazu bildet man die Quadratsumme der Residuen (Residual Sum of Squares),

$$SS_{\text{Res}} = \sum_{i=1}^n (y_i - \hat{y}_i)^2. \quad (2.14)$$

Eine Schätzung für die Varianz σ^2 ergibt sich aus (Montgomery, S.23)

$$\hat{\sigma}^2 = \frac{SS_{\text{Res}}}{n-2} = MS_{\text{Res}}. \quad (2.15)$$

Im Nenner steht dabei $n-2$, die Anzahl der Freiheitsgrade von SS_{Res} . Zwei Freiheitsgrade werden abgezählt für die beiden Bestimmungsgleichungen von $\hat{\beta}_0$ und $\hat{\beta}_1$. Anschaulich kann man sich überlegen, dass bei nur zwei Punkten ($n=2$) die Regressionsgerade durch diese Punkte geht und daher keine Freiheitsgrade mehr vorhanden sind. Mit der geschätzten Varianz $\hat{\sigma}^2$ kann man die Standardfehler (standard error) der Regressionskoeffizienten definieren:

$$se(\hat{\beta}_0) = \sqrt{MS_{\text{Res}} \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right)} \quad \text{und} \quad se(\hat{\beta}_1) = \sqrt{\frac{MS_{\text{Res}}}{S_{xx}}}. \quad (2.16)$$

Ein wichtiges Resultat ist der Satz von Gauss-Markov. Nach diesem Satz sind die Regressionskoeffizienten, berechnet mit der Methode der kleinsten Quadrate, nicht nur unverzerrte Schätzwerte der wahren Regressionskoeffizienten, sondern sie sind auch diejenigen mit der kleinsten Varianz aus allen unverzerrten Schätzwerten, die als Linearkombination der y_i gebildet werden können. Die Annahmen, unter denen dieser Satz gilt, sind $E(\varepsilon) = 0$ und $\text{var}(\varepsilon) = \sigma^2$, und die Fehler dürfen nicht korreliert sein. Ein Schätzwert mit minimaler Varianz wird auch als effizient bezeichnet. Diese Eigenschaft ist vor allem im Hinblick auf die Prognose wichtig. Unverzerrtheit und Effizienz sind zwei wichtige Eigenschaften von statistischen Schätzwerten (Pindyck, S.28). Zur Illustration der Begriffe dienen Bild 2-2 und Bild 2-3.

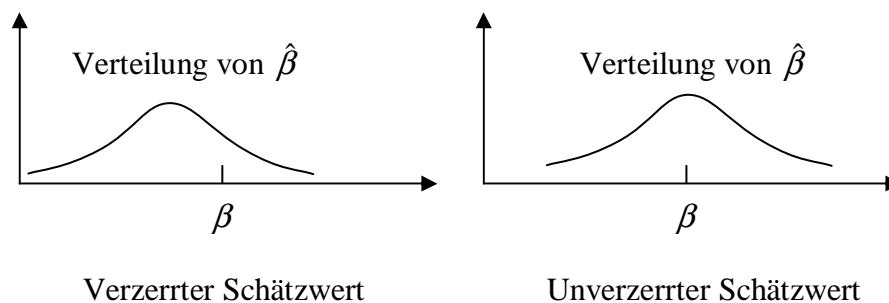


Bild 2-2: Verzernte und unverzernte Schätzwerte

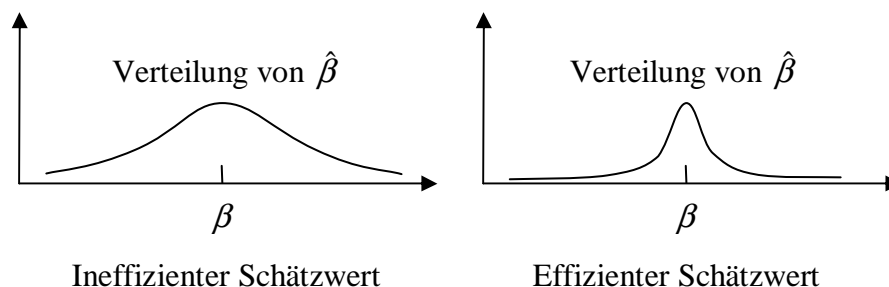


Bild 2-3: Effizienz von Schätzwerten

Oft wird die Regression nur nach den Residuen beurteilt und man macht sich keine Gedanken über die Varianz der Regressionskoeffizienten. Selbst wenn die Annahmen des Satzes von Gauss-Markov erfüllt sind, ist die Varianz nur minimal bezüglich aller unverzerrten Schätzwerte. Bei einer Multikollinearität kann diese Varianz so gross sein, dass man einen verzerrten Schätzwert mit einer kleineren Varianz vorzieht. Das ist genau was man tut, wenn man mit der Principal-Component-Regression oder mit der Ridge-Regression die Multikollinearität beseitigt. Wenn die Annahmen der Methode der kleinsten Quadrate verletzt sind, haben die Regressionskoeffizienten nicht mehr automatisch eine minimale Varianz, wie später im Abschnitt über die Autokorrelation der Fehler erläutert wird.

2.2.2 Varianzanalyse (Analysis of Variance, ANOVA)

Eine einfache Übersicht über die verschiedenen Abweichungen bietet die Varianzanalyse. Wie aus Bild 2-4 hervor geht, setzt sich die totale Abweichung aus der Regressionsabweichung und dem Residuum zusammen.

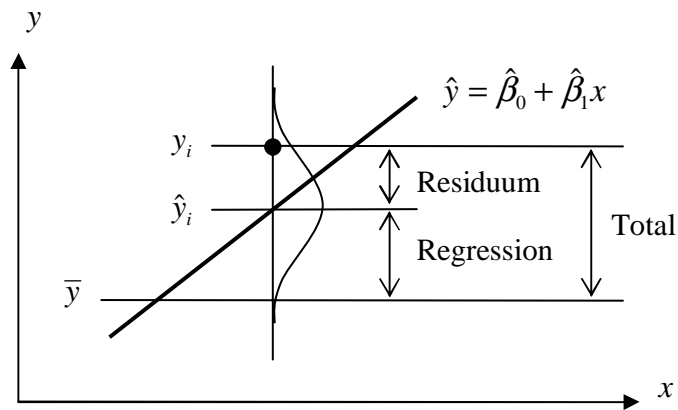


Bild 2-4: Varianzanalyse

Es ist also

$$y_i - \bar{y} = (y_i - \hat{y}_i) + (\hat{y}_i - \bar{y}). \quad (2.17)$$

Von allen drei Differenzen in obiger Gleichung werden Quadratsummen definiert. Die Quadratsumme der Residuen (Residual Sum of Squares) ist

$$SS_{\text{Res}} = \sum_{i=1}^n (y_i - \hat{y}_i)^2. \quad (2.18)$$

Die Quadratsumme der Regressionsabweichung (Regression Sum of Squares) ist

$$SS_{\text{Reg}} = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \quad (2.19)$$

und die Quadratsumme der totalen Abweichung ist

$$SS_{\text{Tot}} = \sum_{i=1}^n (y_i - \bar{y})^2. \quad (2.20)$$

Analog den Abweichungen gilt auch für die Quadratsummen

$$SS_{\text{Tot}} = SS_{\text{Res}} + SS_{\text{Reg}}. \quad (2.21)$$

Die gemischten Terme aus der Quadrierung fallen weg, da die Residuen $y_i - \hat{y}_i$ und die Regressionsabweichungen $\hat{y}_i - \bar{y}$ senkrecht aufeinander stehen (Satz des Pythagoras).

Eine weitere Betrachtung bezieht sich auf die Anzahl der Freiheitsgrade. Die Quadratsumme der Residuen hat

$$df_{\text{Res}} = n - k - 1 \quad (2.22)$$

Freiheitsgrade, wobei n die Anzahl der Beobachtungen ist, und k die Anzahl der unabhängigen Variablen, für die einfache Regression also 1. Die Anzahl der Regressionskoeffizienten $k + 1$ muss subtrahiert werden, da durch die Bestimmungsgleichungen für die Koeffizienten eine entsprechende Anzahl zusätzlicher Bedingungen eingeführt wurde. Die Quadratsumme der Regressionsabweichung hat

$$df_{\text{Reg}} = k \quad (2.23)$$

Freiheitsgrade. Die Konstante wird nicht als Freiheitsgrad gezählt, da sie dadurch bestimmt werden kann, dass die Regressionsgerade durch den Schwerpunkt geht. Die totale Anzahl Freiheitsgrade ist

$$df_{\text{Tot}} = n - 1, \quad (2.24)$$

nämlich gleich der Anzahl der Beobachtungen minus der Bedingung, dass die Regressionsgerade durch den Schwerpunkt geht. Es gilt also auch für die Freiheitsgrade

$$df_{\text{Tot}} = df_{\text{Reg}} + df_{\text{Res}}. \quad (2.25)$$

Weiter können die mittleren Quadratsummen

$$MS_{\text{Res}} = \frac{SS_{\text{Res}}}{n - k - 1} \quad \text{und} \quad MS_{\text{Reg}} = \frac{SS_{\text{Reg}}}{k} \quad (2.26)$$

definiert werden. MS_{Res} wurde bereits früher als Schätzwert für die Varianz der Fehler $\hat{\sigma}^2$ benutzt.

2.2.3 Korrelationskoeffizient

Eine wichtige Grösse zur Beurteilung der Regression ist der Korrelationskoeffizient R . Häufig wird dessen Quadrat, das Bestimmtheitsmass verwendet (coefficient of determination):

$$R^2 = \frac{SS_{\text{Reg}}}{SS_{\text{Tot}}} = 1 - \frac{SS_{\text{Res}}}{SS_{\text{Tot}}}. \quad (2.27)$$

Das Bestimmtheitsmass ist das Verhältnis der durch die Regression erklärten Variation zur totalen Variation. Der Wert ist immer kleiner als Eins, wobei ein Wert in der Nähe von Eins eine gute Übereinstimmung der Regression angibt.

2.2.4 F-Test

Eine andere Grösse zur Beurteilung der Regression ist die F -Kenngrösse

$$F_0 = \frac{SS_{\text{Reg}} / df_{\text{Reg}}}{SS_{\text{Res}} / df_{\text{Res}}} = \frac{SS_{\text{Reg}} / 1}{SS_{\text{Res}} / (n - 2)} = \frac{MS_{\text{Reg}}}{MS_{\text{Res}}} \quad (2.28)$$

wobei MS_{Reg} und MS_{Res} bereits oben definiert wurden. Die F -Kenngrösse ist das Verhältnis der erklärten Varianz zur unerklärten Varianz. Man erwartet ein grosses Verhältnis bei einer guten Übereinstimmung der Regression. Die F -Kenngrösse folgt einer F -Verteilung mit 1 und $n - 2$ Freiheitsgraden und kann für den so genannten F -Test verwendet werden. Die Hypothese, dass $\beta_1 = 0$, dass also kein linearer Zusammenhang zwischen den Variablen besteht, wird verworfen, wenn

$$F_0 > F_{\alpha, 1, n-2}. \quad (2.29)$$

Dabei kann der Wert $F_{\alpha, 1, n-2}$ für ein bestimmtes Signifikanzniveau α in einer Tabelle nachgeschlagen werden. In Computerprogrammen wird meistens ein p -Wert berechnet, sodass $F_0 = F_{p, 1, n-2}$. Die Hypothese wird dann verworfen, wenn $p < \alpha$.

Eine lineare Beziehung zwischen den Variablen kann also nur angenommen werden, wenn $F_0 > F_{\alpha, 1, n-2}$ oder $p < \alpha$.

Der Regressionskoeffizient und die F -Kenngrösse hängen eng zusammen:

$$F_0 = \frac{R^2}{1 - R^2} \frac{n - k - 1}{k}. \quad (2.30)$$

Die F -Kenngrösse hat aber im Gegensatz zum Regressionskoeffizienten eine statistische Bedeutung.

2.3 Signifikanz und Konfidenzintervalle

2.3.1 t-Verteilung

Wie weiter oben erläutert, werden die geschätzten Regressionskoeffizienten $\hat{\beta}_0$ und $\hat{\beta}_1$ als Zufallsgrössen betrachtet. Es wurde auch bereits der Erwartungswert und die Varianz berechnet. Hier soll auch noch die Verteilung bestimmt werden. Aus den Formeln für die Regressionskoeffizienten ist ersichtlich, dass $\hat{\beta}_1$ eine Linearkombination aus den Beobachtungsgrössen y_i ist. Wenn man annimmt, dass die Fehler normal verteilt sind, ist auch $\hat{\beta}_1$ normal verteilt, und zwar mit Mittelwert β_1 und Standardabweichung σ / S_{xx} [Gleichungen (2.12) und (2.13)]. Durch eine Normierung auf den Mittelwert Null und die Varianz Eins erhält man die Testgrösse (test statistic)

$$Z_0 = \frac{\hat{\beta}_1 - \beta_1}{\text{var}(\hat{\beta}_1)} = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\sigma^2 / S_{xx}}}. \quad (2.31)$$

Diese ist definitionsgemäss normal verteilt mit Mittelwert 0 und Varianz 1. Nun ist aber σ^2 nicht bekannt, sondern muss geschätzt werden, nämlich durch $\hat{\sigma}^2 = MS_{\text{Res}}$. Da diese Grösse selbst wieder eine Zufallsvariable ist, ist die Kenngrösse

$$t_0 = \frac{\hat{\beta}_1 - \beta_1}{\text{se}(\hat{\beta}_1)} = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{MS_{\text{Res}} / S_{xx}}}, \quad (2.32)$$

nicht normal verteilt, sondern gehorcht einer t -Verteilung (Student's distribution) mit $n - p$ (hier $n - 2$) Freiheitsgraden. Diese Verteilung ist ähnlich wie die Normalverteilung, nur etwas breiter und abhängig von der Anzahl Freiheitsgraden. Für eine grössere Anzahl von Freiheitsgraden (ab etwa 50) ist der Unterschied praktisch vernachlässigbar. Nur für kleine Stichproben wirkt sich die Unsicherheit in der Schätzung von σ^2 aus, indem die Verteilung etwas breiter ist. Den Vergleich sieht man in Bild 2-5, wo die Dichten der Normalverteilung, sowie der t -Verteilungen mit 2 und 10 Freiheitsgraden geplottet sind.

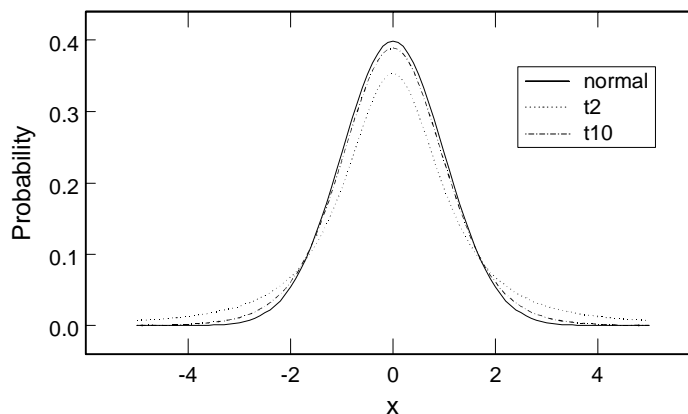


Bild 2-5: t -Verteilung und Normalverteilung

2.3.2 t-Test

Die Frage, ob überhaupt eine lineare Beziehung zwischen den Variablen besteht, wurde bereits im Zusammenhang mit dem F -Test aufgegriffen. Hier soll sie noch einmal aus einer andern Perspektive untersucht werden. Die Frage ist, ob der wahre Koeffizient β_1 nicht Null sein könnte, auch wenn der berechnete Schätzwert $\hat{\beta}_1$ von Null verschieden ist. Die Verteilung von t_0 kann diese Frage nicht direkt beantworten, da sie ja noch vom wahren Koeffizienten β_1 abhängt. Die einzige Möglichkeit ist die, für β_1 einen Wert anzunehmen und zu schauen mit welcher Wahrscheinlichkeit der berechnete Wert $\hat{\beta}_1$ auftreten kann. Ist diese Wahrscheinlichkeit sehr klein, so ist der angenommene Wert kaum richtig. Dies ist die Überlegung eines Hypothesentests. Um also zu überprüfen, ob der geschätzte Wert $\hat{\beta}_1$ signifikant von Null verschieden ist, wird die so genannte Null-Hypothese aufgestellt, dass der wahre Regressionskoeffizient gleich Null ist:

$$H_0 : \beta_1 = 0. \quad (2.33)$$

Trifft die Null-Hypothese zu, dann folgt die Testgrösse

$$t_0 = \frac{\hat{\beta}_1}{\text{se}(\hat{\beta}_1)} = \frac{\hat{\beta}_1}{\sqrt{MS_{\text{Res}} / S_{xx}}} \quad (2.34)$$

einer t -Verteilung mit $n-2$ Freiheitsgraden. Für ein bestimmtes Signifikanzniveau α wird der Grenzwert $t_{\alpha/2, n-2}$ bestimmt. Ist $|t_0| > t_{\alpha/2, n-2}$ wird die Hypothese verworfen. Man beachte, dass t_0 positiv oder negativ sein kann und daher der zweiseitige Test verwendet wird, wie das in Bild 2-6 dargestellt ist.

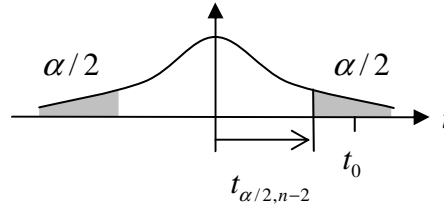


Bild 2-6: Verteilung und Grenzen für t-Test

In Computerprogrammen wird oft ein p -Wert berechnet, sodass $t_{p/2, n-2} = t_0$. Die Hypothese wird dann verworfen, wenn $p < \alpha$.

Ein Regressionskoeffizient ist also signifikant von Null verschieden, wenn $|t_0| > t_{\alpha/2, n-2}$ oder $p < \alpha$.

Die Grösse eines Regressionskoeffizienten selber sagt noch nichts aus darüber, ob er im Modell wichtig ist oder nicht, da er dimensionsbehaftet ist. Die standardisierten Regressionskoeffizienten sind zwar normiert, sagen aber nichts aus über die statistische Relevanz. Ein statistisches Mass für die Wichtigkeit eines Regressionskoeffizienten erhält man, wenn man ihn mit seinen Standardfehler (geschätzte Standardabweichung) normiert. Das ist genau, was der t -Test macht. Die Testgrösse t_0 gibt an, wie viele Standardabweichungen der geschätzte Regressionskoeffizient von Null entfernt ist.

2.3.3 Zusammenhang mit F-Test

Der t -Test und der F -Test beantworten bei der einfachen linearen Regression die gleiche Frage und es gilt der Zusammenhang

$$t_0^2 = \frac{\hat{\beta}_1^2 S_{xx}}{MS_{\text{Res}}} = \frac{\hat{\beta}_1 S_{xy}}{MS_{\text{Res}}} = \frac{MS_{\text{Reg}}}{MS_{\text{Res}}} = F_0. \quad (2.35)$$

Die t -Verteilung mit $n-2$ Freiheitsgraden und die F -Verteilung mit einem und $n-2$ Freiheitsgraden sind identisch. Der Unterschied zeigt sich erst bei der multiplen Regression mit mehreren Freiheitsgraden im Zähler. Der F -Test beantwortet dann die Frage, ob mindestens ein Regressionskoeffizient signifikant von Null verschieden ist, während der t -Test die Signifikanz jedes einzelnen Koeffizienten beurteilt.

2.3.4 Konfidenzintervall für Regressionskoeffizienten

Eine ähnliche Überlegung wie beim t -Test kann für das Konfidenzintervall gemacht werden. Wenn man den wahren Wert β_1 kennen würde, könnte man die Verteilung der geschätzten Werte $\hat{\beta}_1$ angeben. Hier kennt man aber den Schätzwert und will etwas über den wahren Wert aussagen. Die Abweichung des Schätzwertes vom wahren Wert, normiert mit dem Standardfehler,

$$\frac{\hat{\beta}_1 - \beta_1}{\text{se}(\hat{\beta}_1)} \quad (2.36)$$

folgt einer t -Verteilung mit $n-2$ Freiheitsgraden. Für ein gewisses Konfidenzniveau $-\alpha$ liegt diese Grösse im Intervall

$$\left| \frac{\hat{\beta}_1 - \beta_1}{\text{se}(\hat{\beta}_1)} \right| < t_{\alpha/2, n-2}. \quad (2.37)$$

Daraus ergibt sich das Konfidenzintervall

$$\hat{\beta}_1 - t_{\alpha/2, n-2} \sqrt{MS_{\text{Res}} / S_{xx}} \leq \beta_1 \leq \hat{\beta}_1 + t_{\alpha/2, n-2} \sqrt{MS_{\text{Res}} / S_{xx}}. \quad (2.38)$$

Umfasst das Konfidenzintervall den Wert Null, so ist das Gleiche, wie wenn im entsprechenden t -Test die Null-Hypothese akzeptiert wird (also β_1 nicht signifikant von Null verschieden ist). Das Konfidenzintervall und der t -Test führen formal zu gleichen Ergebnissen, wenn auch aus verschiedenen Überlegungen.

2.3.5 Konfidenzintervall für Erwartungswert von y

Die Konfidenzintervalle für die Regressionsparameter sind interessant, wenn man direkt an diesen Parametern interessiert ist. Oft interessiert aber ein Konfidenzintervall für einen Erwartungswert $\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0$ für ein gegebenes x_0 , das nicht unbedingt für die Regression verwendet wurde. Das Intervall ist durch folgende Formel gegeben:

$$\hat{y}_0 - t_{\alpha/2} \sqrt{MS_{\text{Res}} \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right)} \leq E(y | x_0) \leq \hat{y}_0 + t_{\alpha/2} \sqrt{MS_{\text{Res}} \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right)}. \quad (2.39)$$

Wichtig zu sehen ist, dass das Intervall von x_0 abhängt. Am engsten ist es in der Nähe von \bar{x} , während es bei den Extremwerten am weitesten ist. Dies ist schematisch durch die inneren Grenzen in Bild 2-7 dargestellt. Die Regressionsgerade kann sich durch eine vertikale Verschiebung und eine Änderung der Neigung in diesen Grenzen bewegen.

2.3.6 Vorhersageintervall (Prediction interval)

Bei Vorhersagen von einzelnen Werten kommt zur Unsicherheit im Mittelwert auch noch die Streuung der Beobachtungswerte hinzu. Das Intervall ist durch folgende Formel gegeben:

$$\hat{y}_0 - t_{\alpha/2} \sqrt{MS_{\text{Res}} \left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right)} \leq y_0 \leq \hat{y}_0 + t_{\alpha/2} \sqrt{MS_{\text{Res}} \left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right)}. \quad (2.40)$$

Dieses Intervall ist nicht konstant, sondern variiert mit x_0 . Es berücksichtigt ausser der Streuung ε auch die Streuung der Regressionskoeffizienten. In Bild 2-7 ist das Intervall schematisch durch die äusseren Grenzen angegeben.

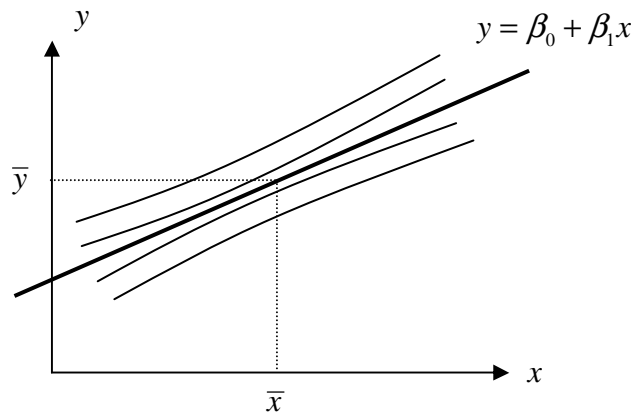


Bild 2-7: Regressionsgerade mit Konfidenzintervallen für Mittelwert und Prognose

2.4 Überprüfung der Annahmen

2.4.1 Normalverteilung der Residuen

Die Annahme der Normalverteilung der Residuen kann durch einen Normalverteilungsplot überprüft werden. Die Residuen e_i , $i = 1, \dots, n$ werden der Grösse nach geordnet und gegen die Funktion

$$x = \Phi^{-1}\left(\left(i - \frac{1}{2}\right)/n\right) \quad (2.41)$$

geplottet, wobei $\Phi(x)$ die Normalverteilungsfunktion darstellt. Sind die Residuen normal verteilt, liegen diese Punkte auf einer Geraden. Das Prinzip ist in Bild 2-8 gezeigt. Man geht von rechts mit der Wahrscheinlichkeit $(i - \frac{1}{2})/n$ ins Diagramm und findet den Wert $\Phi^{-1}\left((i - \frac{1}{2})/n\right)$ auf der x -Achse. Da $\Phi(x)$ auf Mittelwert Null und Standardabweichung Eins normiert ist, ist auch die x -Achse so normiert. Der Fehler selber wird auf der y -Achse aufgetragen.

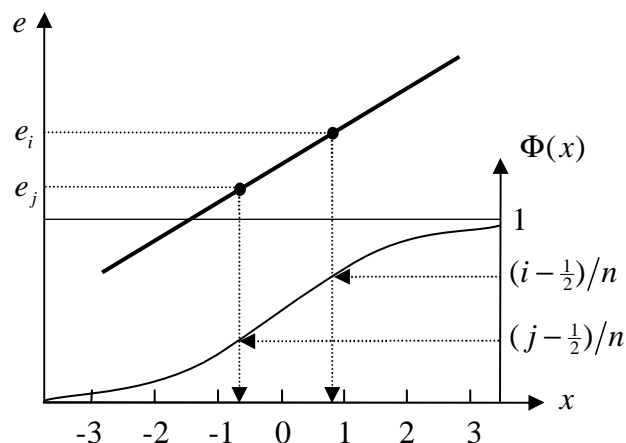


Bild 2-8: Normalverteilung der Residuen

Es ist zu beachten, dass die Annahme einer Normalverteilung nur im Zusammenhang mit statistischen Interpretationen benützt wurde. Die Methode der kleinsten Quadrate kann auch ohne diese Annahme durchgeführt werden, aber die Regressionskoeffizienten sind dann statistisch gesehen nicht mehr unbedingt die besten Schätzwerte. Die t -Testgrösse kann nicht mehr für die Signifikanz benützt werden, aber als Mass des Fehlerzuwachses ist sie immer noch gültig. Daher ist eine geringe Abweichung von der Normalverteilung in der Praxis nicht relevant.

2.4.2 Heteroskedastizität und Autokorrelation

Eine wichtige Annahme bei der Methode der kleinsten Quadrate ist die, dass der Fehler konstante Varianz hat und unkorreliert ist. Fälle, wo diese Annahme verletzt ist, sind in Bild 2-9 und Bild 2-10 gezeigt (Pindyck, S. 60).

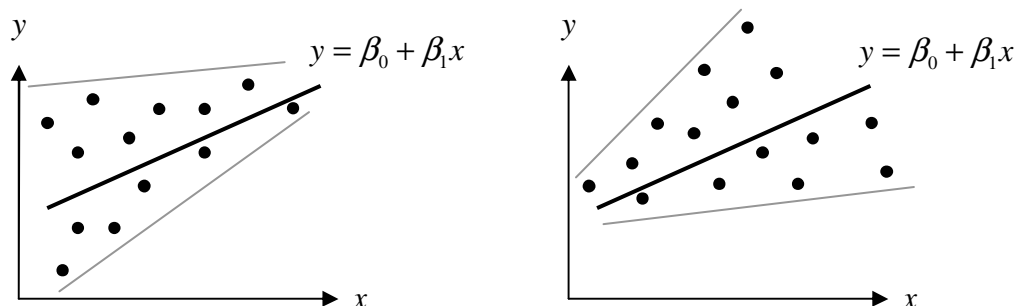


Bild 2-9: Heteroskedastizität

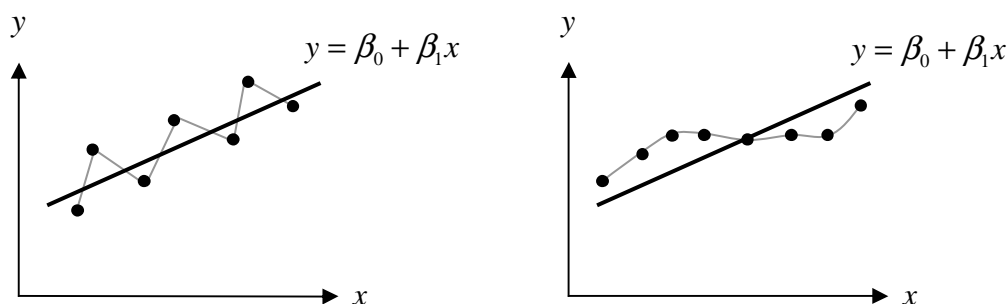


Bild 2-10: Autokorrelation (links negative Korrelation, rechts positive Korrelation)

Eine variable Varianz der Fehler wird als Heteroskedastizität bezeichnet. Obwohl es auch formale Tests dafür gibt, beschränkt sich die vorliegende Studie auf die visuelle Beurteilung der Residuen in Abhängigkeit von \hat{y} .

Wichtiger bei Zeitreihen ist die Autokorrelation der Fehler (auch als serielle Korrelation bezeichnet). Diese kann visuell mit Hilfe des Korrelogramms beurteilt werden. Das Korrelogramm ist die normierte Autokorrelationsfunktion der Residuen:

$$f_t = \frac{\sum_{i=t+1}^n e_i e_{i-t}}{\sum_{i=1}^n e_i^2} \quad t = 0, 1, \dots \quad (2.42)$$

Die Funktion wird hier in der diskreten Form verwendet, wobei vorausgesetzt wird, dass die Werte mit einem konstanten Zeitschritt verteilt sind. Das Argument t ist die zeitliche Verschiebung (time lag). Für $t=0$ wird der Wert der Funktion gleich Eins. Wenn keine Korrelation vorliegt, sind die andern Werte zufällig um die Nulllinie verteilt mit einer Varianz von $1/n$. Bei einer Normalverteilung liegen 95% der Punkte innerhalb der Grenzen $\pm 2/\sqrt{n}$. Üblicherweise werden zwei horizontale Linien bei diesen Grenzen eingezeichnet, die helfen, das Korrelogramm zu beurteilen (Harvey 1992, S.52). Eine schematische Darstellung sieht man in Bild 2-11.

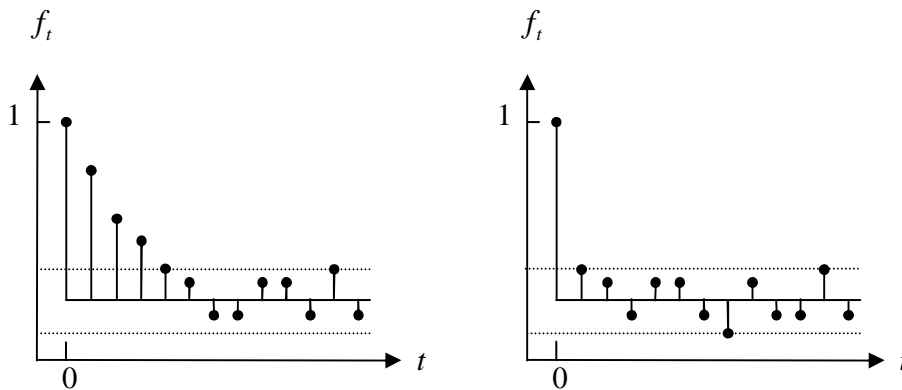


Bild 2-11: Schematische Darstellung Korrelogramm, links mit (positiver) Autokorrelation, rechts ohne Autokorrelation

Formal kann die Autokorrelation mit dem Durbin-Watson-Test nachgewiesen werden. Die Kenngrösse ist

$$d = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n e_i^2}. \quad (2.43)$$

Der Durbin-Watson-Test kann allerdings nur eine bestimmte, häufig vorkommende Art der Korrelation feststellen, nämlich die Korrelation erster Ordnung. Das Modell für den Fehler ist

$$\varepsilon_t = \rho \varepsilon_{t-1} + u_t. \quad (2.44)$$

Der Fehler ε_t zur Zeit t setzt sich zusammen aus einem Anteil ρ des vorangehenden Fehlers und einem neuen Fehler u_t mit Mittelwert Null und Varianz σ_u^2 . Der Faktor ρ ($|\rho| < 1$) ist der Autokorrelationsparameter. Die Kenngrösse d kann einen Wert zwischen 0 und 4 annehmen. Der Zusammenhang zum Autokorrelationsparameter ist

$$d \approx 2(1 - \rho). \quad (2.45)$$

Null bedeutet eine positive Korrelation, 2 keine Korrelation und 4 eine negative Korrelation. Um den Test durchzuführen, muss die Verteilung von d bekannt sein. Diese hängt aber vom Regressionsmodell ab. Als praktische Lösung haben Durbin und Watson (1950, 1951) untere und obere Grenzwerte d_L und d_U in Abhängigkeit der Anzahl

Beobachtungen n und der Anzahl Regressionsparameter angegeben (für ein bestimmtes Signifikanzniveau). Die Bereiche mit positiver, negativer und ohne Korrelation sind in Bild 2-12 angegeben. In den Zwischenbereichen kann mit den Grenzwerten allein keine Entscheidung getroffen werden. Für diese Fälle geben Durbin und Watson weitere Verfahren an, die aber hier nicht weiter besprochen werden sollen. Die Werte d_L und d_U können aus Tabellen herausgelesen werden. Um einen generellen Eindruck zu vermitteln, sind in Tabelle 2-1 ein paar Werte angegeben. Für grosse Stichproben folgen die Kenngrössen ungefähr einer Beta-Verteilung. Formeln für den Mittelwert und die Varianz sind in Durbin und Watson (1950, S. 427) angegeben.

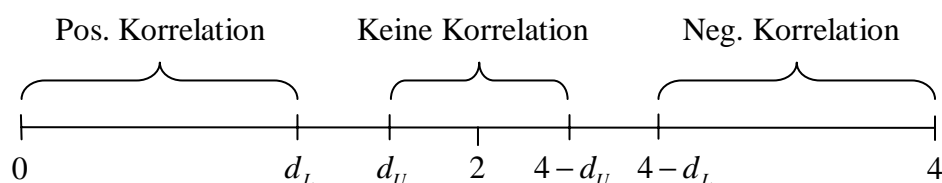


Bild 2-12: Bereiche der Durbin-Watson-Kenngrösse

Tabelle 2-1: Grenzwerte für die Durbin-Watson-Verteilung für ausgewählte Parameter (5% Signifikanzniveau)

	5 Regressoren		10 Regressoren	
n	d_L	d_U	d_L	d_U
50	1.335	1.771	1.110	2.044
100	1.571	1.780	1.462	1.898
150	1.665	1.802	1.593	1.877
200	1.718	1.820	1.665	1.874

Was die Autokorrelation für Konsequenzen hat für die Regression, und wie sie behandelt werden kann, ist in Kapitel 4 beschrieben.

3 Multiple Regression

Bei der multiplen Regression werden mehrere unabhängige Variablen betrachtet anstatt nur einer einzelnen. Die unabhängigen Variablen werden oft auch als Regressoren bezeichnet. Der Begriff ist etwas allgemeiner und schliesst transformierte Variablen mit ein. Viele Herleitungen der einfachen Regression können für die multiple Regression verallgemeinert werden und werden hier nur als Resultate wiedergegeben. Einige Konzepte wie der F -Test und der t -Test bekommen eine etwas andere Bedeutung als bei der einfachen Regression. Ein neues Problem, das nur bei der multiplen Regression auftreten kann, ist die Multikollinearität. Ein wichtiges Thema ist schliesslich die Modellbildung und Variablenauswahl aufgrund der statistischen Kenngrössen.

3.1 Methode der kleinsten Quadrate

Für die multiple Regression wird am besten die Darstellung mit Matrizen und Vektoren verwendet, da sie eine kompakte und übersichtliche Darstellung erlaubt. Es werden folgende Matrizen und Vektoren benützt:

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1k} \\ 1 & x_{21} & & x_{2k} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \cdots & x_{nk} \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} \quad \text{und} \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}. \quad (3.1)$$

Die Matrix \mathbf{X} hat die Dimension $n \times (k+1)$, wobei n die Anzahl der Beobachtungen und k die Anzahl der Regressoren bezeichnet.

Wie bei der einfachen Regression wird \mathbf{y} als Zufallsvariable betrachtet, während die Werte in \mathbf{X} als gegeben (ohne Messfehler) angenommen werden. Die Zufallsvariable \mathbf{y} wird beschrieben durch

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}. \quad (3.2)$$

Dabei beschreibt die allgemeine lineare Funktion $\mathbf{X}\boldsymbol{\beta}$ den Mittelwert von \mathbf{y} und die Zufallsvariable $\boldsymbol{\varepsilon}$ die Verteilung der Fehler.

Für die Methode der kleinsten Quadrate bildet man die Quadratsumme

$$S = \boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} = (\mathbf{y}^T - \boldsymbol{\beta}^T \mathbf{X}^T)(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}). \quad (3.3)$$

Diese Funktion wird minimiert wenn die Normalgleichungen

$$\mathbf{X}^T \mathbf{X} \boldsymbol{\beta} = \mathbf{X}^T \mathbf{y} \quad (3.4)$$

erfüllt sind. Daraus erhält man die Gleichung für die Koeffizienten $\hat{\boldsymbol{\beta}}$,

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}. \quad (3.5)$$

Der Schätzwert für \mathbf{y} ist

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} \quad (3.6)$$

und die Residuen sind

$$\mathbf{e} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{y} - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} = (\mathbf{I} - \mathbf{H})\mathbf{y}, \quad (3.7)$$

mit der Matrix $\mathbf{H} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$.

3.2 Varianzanalyse

Die Betrachtungen über die Varianzanalyse und den Korrelationskoeffizienten bei der einfachen Regression sind auch für die multiple Regression gültig, da sie nur die abhängige Variable \mathbf{y} betreffen.

3.2.1 Quadratsummen

Die Quadratsumme der Residuen (Residual Sum of Squares) ist

$$SS_{\text{Res}} = \sum_{i=1}^n (y_i - \hat{y}_i)^2. \quad (3.8)$$

Die Quadratsumme der Regressionsabweichung (Regression Sum of Squares) ist

$$SS_{\text{Reg}} = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \quad (3.9)$$

und die Quadratsumme der totalen Abweichung ist

$$SS_{\text{Tot}} = \sum_{i=1}^n (y_i - \bar{y})^2. \quad (3.10)$$

Auch hier gilt die Addition der Quadratsummen

$$SS_{\text{Res}} + SS_{\text{Reg}} = SS_{\text{Tot}}. \quad (3.11)$$

Der Erwartungswert für die Regressionskoeffizienten ist

$$E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}. \quad (3.12)$$

Anstelle der Varianz eines einzelnen Regressionskoeffizienten tritt die Varianz-Kovarianz-Matrix

$$\text{var}(\hat{\boldsymbol{\beta}}) = \sigma^2 (\mathbf{X}^T\mathbf{X})^{-1} \quad (3.13)$$

Die Varianz der Fehler σ^2 wird geschätzt durch

$$\hat{\sigma}^2 = MS_{\text{Res}} = \frac{SS_{\text{Res}}}{n - k - 1}. \quad (3.14)$$

Dabei ist $n - k - 1$ die Anzahl der Freiheitsgrade bei n Beobachtungen und k Regressoren.

3.2.2 Korrelationskoeffizient

Die Anpassung der Regression kann wie bei der einfachen Regression mit dem multiplen Korrelationskoeffizienten oder mit dessen Quadrat, dem Bestimmtheitsmass (Coefficient of multiple determination) beurteilt werden. Diese Grösse ist definiert als

$$R^2 = \frac{SS_{\text{Reg}}}{SS_{\text{Tot}}} = 1 - \frac{SS_{\text{Res}}}{SS_{\text{Tot}}}. \quad (3.15)$$

Je besser die Anpassung bei der Regression ist, umso grösser wird R^2 . Der maximale Wert von Eins wird nur erreicht, wenn die Residuen Null sind.

Der Korrelationskoeffizient wird oft verwendet, um verschiedene Modelle miteinander zu vergleichen. Da jeder zusätzliche, auch unwesentliche Regressor eine Vergrösserung des Korrelationskoeffizienten bringt, ist diese Beurteilung aber nicht ganz objektiv. Daher wird oft der angepasste (adjusted) Korrelationskoeffizient benützt, der zusätzlich die Anzahl der Regressoren berücksichtigt. Die Definition lautet:

$$R_{\text{adj}}^2 = 1 - \frac{SS_{\text{Res}} / (n - k - 1)}{SS_{\text{Tot}} / (n - 1)}. \quad (3.16)$$

Es hat sich jedoch gezeigt, dass der angepasste Korrelationskoeffizient gegenüber dem gewöhnlichen Korrelationskoeffizienten kaum einen Vorteil bringt, wenn die Anzahl der Regressoren k gegenüber der Anzahl Beobachtungen n klein ist. Der Koeffizient R_{adj} kann auch geschrieben werden als

$$R_{\text{adj}}^2 = 1 - \frac{MS_{\text{Res}}}{SS_{\text{Tot}} / (n - 1)}. \quad (3.17)$$

Als Vergleichskriterium sind daher R_{adj} und MS_{Res} äquivalent, d.h. ein Modell mit einem kleineren MS_{Res} wird automatisch auch einen grösseren Wert für R_{adj} anzeigen.

Weitere Kriterien für die Qualität der Regression, die hier nicht untersucht wurden, sind Mallows' C_p -Wert und PRESS (Montgomery, S. 299 und S. 301).

3.2.3 Globaler F-Test

Zusammen mit der Varianzanalyse wird oft der globale F -Test angegeben. Aus der Testgrösse

$$F_0 = \frac{SS_{\text{Reg}} / df_{\text{Reg}}}{SS_{\text{Res}} / df_{\text{Res}}} = \frac{SS_{\text{Reg}} / k}{SS_{\text{Res}} / (n - k - 1)} = \frac{MS_{\text{Reg}}}{MS_{\text{Res}}} \quad (3.18)$$

ist ersichtlich, ob überhaupt eine statistisch relevante lineare Beziehung vorhanden ist. Wenn der Wert für F_0 gross ist, kann eine lineare Beziehung angenommen werden. Genauer gesagt wird die Null-Hypothese getestet, dass alle Regressionskoeffizienten Null sind. Die Hypothese wird verworfen, wenn für ein Signifikanzniveau α

$$F_0 > F_{\alpha, k, n-k-1}. \quad (3.19)$$

Der Wert $F_{\alpha, k, n-k-1}$ kann in einer Tabelle nachgeschlagen werden. In einem Computerprogramm wird meistens der zugehörige p -Wert berechnet. Die Hypothese wird dann verworfen, wenn $p < \alpha$. Die Verwerfung der Null-Hypothese bedeutet die Annahme der

Gegenhypothese, dass mindestens ein Regressionskoeffizient von Null verschieden sein muss. Die Signifikanz der einzelnen Regressionskoeffizienten wird dann mit dem t -Test beurteilt.

Der globale F -Test beurteilt, ob mindestens ein Regressionskoeffizient signifikant von Null verschieden ist. Dies ist der Fall wenn $F_0 > F_{\alpha, k, n-k-1}$ oder $p < \alpha$.

Bei einigermaßen vernünftigen Modellen ist der globale F -Test praktisch immer erfüllt.

3.3 Partieller F-Test, partieller Korrelationskoeffizient

3.3.1 Partieller F-Test bei Elimination

Die F -Kenngrösse kann auch noch anders interpretiert werden, nämlich über die Zunahme der Quadratsumme der Residuen, wenn man alle Regressoren weglassen würde. Es ist einfach zu sehen, dass dann $\hat{y}_i = \bar{y}$ wäre. Die Zunahme der Quadratsumme der Residuen wäre dann gerade $\Delta SS_{\text{Res}} = SS_{\text{Reg}}$, die Grösse, die im Zähler der Definition der F -Testgrösse steht. Diese Definition lässt sich erweitern für den Fall, wo nur ein Teil der Regressoren weggelassen wird. Man spricht dann von einem partiellen F -Test. Wenn durch das Weglassen von r Regressoren die Quadratsumme der Residuen um ΔSS_{Res} zunimmt, lautet die entsprechende F -Testgrösse

$$F_0 = \frac{\Delta SS_{\text{Res}} / r}{MS_{\text{Res}}}. \quad (3.20)$$

Weglassen von r Regressoren bedeutet in diesem Fall eine neue Berechnung der Regressionskoeffizienten mit dem reduzierten Modell. Der Wert MS_{Res} bezieht sich auf das ursprüngliche Modell.

Praktisch interessant ist der Fall, wenn man einen einzelnen Regressor weglässt. Es kann gezeigt werden, dass das Weglassen des Regressors j im reduzierten Modell zu einem Zuwachs der Quadratsumme der Residuen von

$$\Delta SS_{\text{Res}} = \frac{\hat{\beta}_j^2}{C_{jj}} \quad (3.21)$$

führt, wobei C_{jj} den Diagonalterm (j, j) der Matrix $(\mathbf{X}^T \mathbf{X})^{-1}$ bezeichnet. Die Herleitung kann elegant als Regressionsproblem mit der Nebenbedingung $\hat{\beta}_j = 0$ formuliert und mit einem Lagrange-Multiplikator gelöst werden (siehe Anhang). Damit wird die entsprechende partielle F -Kenngrösse für einen einzelnen Regressionskoeffizienten

$$F_0 = \frac{\hat{\beta}_j^2}{MS_{\text{Res}} C_{jj}}. \quad (3.22)$$

Die partielle F -Kenngrösse gibt also an, wie gross der Zuwachs der Quadratsumme der Residuen ist, wenn ein Regressor eliminiert wird. Die gleiche Information liefert auch die t -Kenngrösse. Wie weiter unten gezeigt wird, ist die partielle F -Kenngrösse für einen Regressor gerade gleich dem Quadrat der t -Kenngrösse.

3.3.2 Partieller F-Test bei Einfügen, partieller Korrelationskoeffizient

Möchte man zu einem bestehenden Modell einen neuen Regressor hinzufügen und kann man aus mehreren auswählen, so stellt sich die Frage, welcher die Quadratsumme am stärksten reduzieren kann. Wie im Anhang gezeigt wird, ist die Abnahme der Quadratsumme der Residuen beim Einfügen eines neuen Regressors \mathbf{z}

$$\Delta SS_{\text{Res}} = \frac{(\mathbf{y}^T \mathbf{e}_z)^2}{\mathbf{z}^T \mathbf{e}_z} = \frac{(\mathbf{e}_y^T \mathbf{e}_z)^2}{\mathbf{e}_z^T \mathbf{e}_z}. \quad (3.23)$$

Dabei ist

$$\mathbf{e}_y = \mathbf{y} - \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \quad (3.24)$$

der Vektor der Residuen und analog

$$\mathbf{e}_z = \mathbf{z} - \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{z} \quad (3.25)$$

der Vektor der Residuen, wenn man die Regression mit \mathbf{z} anstatt mit \mathbf{y} durchführt. Eine entsprechende F -Kenngrösse bezogen auf das Modell mit den neuen Regressor ist dann

$$F_0 = \frac{\Delta SS_{\text{Res}}}{(SS_{\text{Res}} - \Delta SS_{\text{Res}})/(n - k - 2)} \quad (3.26)$$

wobei sich SS_{Res} und k auf das ursprüngliche Modell vor dem Einfügen des Regressors beziehen.

Ein ähnliches Resultat folgt aus der Überlegung mit dem so genannten partiellen Korrelationskoeffizienten (Goldberger 1966, S. 198). Die Idee ist, die Residuen \mathbf{e}_y durch den zusätzlichen Regressor zu modellieren. Man ist aber nur an dem Teil interessiert, der nicht schon durch die alten Regressoren modelliert werden kann. Daher zieht man auch vom neuen Regressor den Anteil ab, der durch die alten Regressoren ausgedrückt werden kann, d.h. man betrachtet die Residuen \mathbf{e}_z . Der partielle Korrelationskoeffizient ist dann der gewöhnliche Korrelationskoeffizient zwischen den Residuen \mathbf{e}_y und \mathbf{e}_z . Das Quadrat des Korrelationskoeffizienten ist also

$$R_{y,z}^2 = \frac{(\mathbf{e}_y^T \mathbf{e}_z)^2}{(\mathbf{e}_z^T \mathbf{e}_z)(\mathbf{e}_y^T \mathbf{e}_y)}. \quad (3.27)$$

Beachtet man dass $SS_{\text{Res}} = \mathbf{e}_y^T \mathbf{e}_y$, findet man durch Vergleich mit (3.23), dass die obige Formel auch ausgedrückt werden kann als

$$R_{y,z}^2 = \frac{\Delta SS_{\text{Res}}}{SS_{\text{Res}}}, \quad (3.28)$$

wobei sich SS_{Res} wieder auf das ursprüngliche Modell vor dem Einfügen des Regressors bezieht. Der partielle Korrelationskoeffizient lässt sich mit Hilfe von (3.26) auch ausdrücken als

$$R_{y,z}^2 = \frac{F_0}{F_0 + (n - k - 2)}. \quad (3.29)$$

Der partielle Korrelationskoeffizient liefert also im Wesentlichen die gleiche Information wie die partielle F -Kenngrösse oder damit auch wie die t -Kenngrösse, die ein Regressor nach dem Einfügen bekommen würde.

3.4 Signifikanz und Konfidenzintervalle

3.4.1 t-Test

Wie bei der einfachen Regression kann auch bei der multiplen Regression ein t -Test für einzelne Regressionskoeffizienten durchgeführt werden. Der Standardfehler eines Regressionskoeffizienten wird aus dessen Varianz [Gleichung (3.13)] und dem Schätzwert $\hat{\sigma}^2 = MS_{\text{Res}}$ bestimmt:

$$\text{se}(\hat{\beta}_j) = \sqrt{\hat{\sigma}^2 C_{jj}}. \quad (3.30)$$

Die t -Kenngrösse ist damit

$$t_0 = \frac{\hat{\beta}_j}{\text{se}(\hat{\beta}_j)} = \frac{\hat{\beta}_j}{\sqrt{MS_{\text{Res}} C_{jj}}}. \quad (3.31)$$

Dieser Wert und der entsprechende t -Test gilt nur für die aktuelle Kombination von Regressoren. Wenn das Modell ändert, ändern sich auch die t -Kenngrössen.

Wie schon bei der einfachen Regression gibt es einen Zusammenhang zwischen der t -Kenngrösse und der partiellen F -Kenngrösse für einen einzelnen Regressor, nämlich $t_0^2 = F_0$. Die t -Kenngrösse (wie die F -Kenngrösse) hat somit nicht nur die statistische Bedeutung, die etwas über die Signifikanz der Regressionskoeffizienten aussagt, sondern ist auch ein direktes Mass dafür, wie stark die Quadratsumme der Residuen zunimmt, wenn ein einzelner Regressor aus dem Modell eliminiert wird. Er ist daher ein wichtiges Mittel bei der Variablenauswahl und der Modellbildung.

Der t -Test und der partielle F -Test für einen Regressor sind äquivalent. Für die Testgrössen gilt $t_0^2 = F_0$. Ein einzelner Regressionskoeffizient ist signifikant, wenn die Testgrösse über dem entsprechenden kritischen Wert liegt, oder wenn $p < \alpha$. Der Test gilt nur für die aktuelle Kombination von Regressoren, d.h. nur den nächsten Eliminatioonsschritt.

3.4.2 Konfidenz- und Prognoseintervalle

Wie bei der einfachen Regression können Konfidenzintervalle für die Regressionsparameter und für die Prognose angegeben werden. Das $100(1 - \alpha)\%$ Konfidenzintervall für einen Regressionsparameter β_j ist

$$\hat{\beta}_j - t_{\alpha/2, n-k-1} \sqrt{\hat{\sigma}^2 C_{jj}} \leq \beta_j \leq \hat{\beta}_j + t_{\alpha/2, n-k-1} \sqrt{\hat{\sigma}^2 C_{jj}}. \quad (3.32)$$

Das Konfidenzintervall für einen einzelnen Wert y_0 in der Zukunft muss sowohl die Streuung des Mittelwertes wie auch die Streuung der y -Werte um den Mittelwert berücksichtigen. Das $100(1 - \alpha)\%$ Prognoseintervall für y_0 an einem Punkt \mathbf{x}_0 (der nicht unbedingt bei der Regression benutzt wurde) ist

$$\begin{aligned} \hat{y}_0 - t_{\alpha/2, n-p} \sqrt{\hat{\sigma}^2 (1 + \mathbf{x}_0^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_0)} \\ \leq y_0 \leq \hat{y}_0 + t_{\alpha/2, n-p} \sqrt{\hat{\sigma}^2 (1 + \mathbf{x}_0^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_0)} \end{aligned} \quad (3.33)$$

Dabei ist $\hat{y}_0 = \mathbf{x}_0^T \hat{\boldsymbol{\beta}}$ der Erwartungswert an der Stelle \mathbf{x}_0 . Dieses Intervall ist allgemeiner als das oft in der Praxis verwendete Intervall von $\pm m\hat{\sigma}$ (Faktor mal Standardabweichung), das den Einfluss der Streuung der Regressionsparameter vernachlässigt. Die Streuung der Regressionsparameter wird durch den Ausdruck $\mathbf{x}_0^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_0$ berücksichtigt. Der Ausdruck $\mathbf{x}_0^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_0$ wird vor allem wichtig, wenn der Punkt \mathbf{x}_0 ausserhalb des Regressionsbereichs liegt, wie das im nächsten Abschnitt beschrieben ist.

3.4.3 Extrapolation

Ein Regressionsmodell liefert nur dann eine zuverlässige Prognose, wenn die Regressionsvariablen im ursprünglichen Regressionsbereich liegen. Bei der multiplen Regression ist dieser Bereich nicht einfach durch die Bereiche der einzelnen Regressoren gegeben, sondern durch deren Kombination (Punkte in einem mehrdimensionalen Raum). Der Interpolationsbereich ist dann ein Gebiet, das diese Kombinationspunkte umfasst. Für zwei Regressoren kann der Sachverhalt aus Bild 3-1 ersehen werden. Da eine Extrapolation nicht ohne weiteres ersichtlich ist, spricht man auch von versteckter Extrapolation.

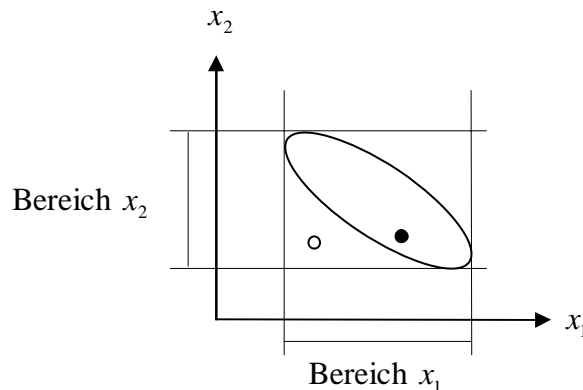


Bild 3-1: Versteckte Extrapolation, Punkte mit: ○ Extrapolation, • Interpolation

Eine einfache Abschätzung durch ein (mehrdimensionales) Ellipsoid kann über die Matrix

$$\mathbf{H} = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \quad (3.34)$$

erfolgen (Montgomery, S. 110). Ein Ellipsoid, das den Bereich der Interpolation umfasst, wird beschrieben durch

$$\mathbf{x}^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x} \leq h_{\max}, \quad (3.35)$$

wobei \mathbf{x} ein allgemeiner Punkt ist. Die Grenze h_{\max} wird aus dem maximalen Diagonalewert h_{ii} der Matrix \mathbf{H} bestimmt. Für einen Prognosepunkt \mathbf{x}_0 bestimmt man dann

$$h_{00} = \mathbf{x}_0^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_0 \quad (3.36)$$

und vergleicht den Wert mit h_{\max} . Ist $h_{00} > h_{\max}$, ist der Punkt \mathbf{x}_0 ein Extrapolationspunkt und man kann nicht unbedingt eine gute Prognose erwarten. Dieser Sachverhalt wird sich auch dadurch zeigen, dass das Prognoseintervall [Gleichung (3.33)] vergrößert wird, da dort der Ausdruck für h_{00} erscheint.

3.5 Multikollinearität

Ein wichtiges Problem, das bei der multiplen Regression auftreten kann, ist die so genannte Multikollinearität. Sind zwei oder mehrere Regressoren von einander linear abhängig, so wird die Matrix $\mathbf{X}^T\mathbf{X}$ singulär. Eine exakte lineare Abhängigkeit wird kaum vorkommen, aber auch eine annähernd lineare Abhängigkeit führt zu verschiedenen Problemen. Eine solche annähernd lineare Abhängigkeit wird hier als Multikollinearität bezeichnet (Kollinearität bei zwei Regressoren). Die Matrix $\mathbf{X}^T\mathbf{X}$ ist schlecht konditioniert und damit kann die Lösung je nach verwendeter Methode numerisch ungenau werden. Eine schlecht konditionierte Matrix bedeutet auch, dass eine kleine Änderung in den Eingabewerten einen grossen Einfluss auf die Lösung hat, ein Verhalten, dass insbesondere für eine Prognose nicht erwünscht ist.

3.5.1 Variable Inflation Factor, VIF

Der wichtigste Parameter, der eine Multikollinearität anzeigt, ist der Variance Inflation Factor, VIF, definiert als

$$\text{VIF}_j = C_{jj} S_{jj}. \quad (3.37)$$

Der VIF ist also im Wesentlichen der Diagonalterm der inversen Matrix $\mathbf{C} = (\mathbf{X}^T\mathbf{X})^{-1}$. Der Skalierungsfaktor S_{jj} ist gegeben durch

$$S_{jj} = \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2. \quad (3.38)$$

Ein grosser VIF-Wert und damit ein grosser Diagonalterm in $(\mathbf{X}^T\mathbf{X})^{-1}$ zeigt direkt, dass eine kleine Variation in den Eingabegrössen die Regressionskoeffizienten stark beeinflusst. Eine andere Interpretation erfolgt über die Varianz der Regressionskoeffizienten:

$$\text{VIF}_j = \frac{\text{var}(\hat{\beta}_j)}{\sigma^2} S_{jj}. \quad (3.39)$$

Ein grosser VIF-Wert bedeutet eine grosse Varianz des entsprechenden Regressionskoeffizienten, daher auch der Name (Variance Inflation Factor). Eine weitere interessante Interpretation beruht auf der Tatsache, dass

$$\text{VIF}_j = \frac{1}{1 - R_j^2}. \quad (3.40)$$

Dabei ist R_j^2 das Bestimmtheitsmass aus einer Regression von \mathbf{x}_j gegenüber allen anderen Regressoren (siehe Anhang). Ein Wert nahe bei Eins ergibt eine hohe Korrelation und zeigt, dass \mathbf{x}_j durch eine lineare Kombination der übrigen Regressoren approximiert werden kann, dass also eine Multikollinearität vorhanden ist. Für ein R_j^2 nahe bei Eins wird der VIF-Wert entsprechend gross.

3.5.2 Auswirkungen der Multikollinearität

Regressionskoeffizienten, die an einer Multikollinearität beteiligt sind, werden nur als Kombination richtig berechnet. Wegen der gegenseitigen Abhängigkeit kann praktisch die gleiche Regression mit verschiedenen Kombinationen erreicht werden. Die einzelnen Koeffizienten haben in dem Sinne keine Aussagekraft. Ist man an den Regressionskoeffizienten direkt interessiert, so muss die Multikollinearität zuerst beseitigt werden. Methoden sind im nächsten Kapitel beschrieben.

Die t -Kenngrößen für Regressoren, die an einer Multikollinearität beteiligt sind, gelten nur für den nächsten Eliminationsschritt, nicht aber für das Gesamtmodell. Im Allgemeinen sind die t -Kenngrößen zu klein und die Bedeutung der Regressoren wird unterschätzt. Eine Multikollinearität wirkt sich bei der Variablen-Selektion ungünstig aus, da bei einer Auswahl aufgrund der t -Kenngrößen oft zuerst die falschen Regressoren eliminiert werden. Auf jeden Fall sollte die Elimination schrittweise erfolgen, damit sich die t -Kenngrößen nach jedem Schritt wieder neu einstellen können.

Für die Prognose ist die Multikollinearität nicht unbedingt nachteilig, solange gewährleistet ist, dass die gegenseitigen Abhängigkeiten der Regressoren auch in der Zukunft gegeben sind. Ein solches Verhalten ist oft bei gemessenen Temperaturen zu beobachten. Ein sehr ungünstiger Fall ist aber der, wenn zwei ähnliche Regressoren sich gegenseitig fast aufheben. Die Differenz ist meist eher zufällig, und man kann nicht damit rechnen, dass das Verhalten für die Prognose gleich ist wie bei der Regression.

Die Bedeutung und der Einfluss der Multikollinearität sind nicht immer klar zu erkennen. Grundsätzlich sollte man sie daher vermeiden, indem man versucht möglichst unabhängige Regressoren zu benutzen. Bei Transformationen von gemessenen Variablen sollte man nach Möglichkeit orthogonale Funktionen wählen. Dieser Punkt wird ausführlicher in Kapitel 5 erläutert. Im nächsten Kapitel werden Methoden beschrieben, mit denen die Multikollinearität beseitigt werden kann.

3.6 Variablenauswahl

Die Statistik ist ein wichtiges Werkzeug beim Aufstellen eines geeigneten Regressionsmodells. Obwohl man sich in erster Linie auf die Physik und die Ingenieur Erfahrung abstützen sollte, kommen doch meistens zu viele Regressoren als Kandidaten in Frage, von denen man die wichtigsten auswählen muss. Für die Variablenauswahl gibt es grundsätzlich zwei Strategien: Man kann von einem Modell mit vielen Regressoren ausgehen und die unwichtigen eliminieren, oder man kann mit einem kleinen Modell beginnen und die wichtigsten Regressoren einfügen.

3.6.1 Elimination

Die Auswahl der Regressoren aus einem grossen Modell erfolgt hauptsächlich auf Grund der t -Testgrößen. Wie schon früher beschrieben, kann diese Grösse auf zwei verschiedene Arten interpretiert werden. Statistisch gesehen sagt sie aus, ob ein Regressionskoeffizient signifikant von Null verschieden ist. Dazu wird der aus der t -Verteilung (Student-Verteilung) berechnete p -Wert verwendet. Ist dieser grösser als ein gewählter Wert (oft 5% oder 1%), ist der entsprechende Regressor nicht signifikant und kann

eliminiert werden. Numerisch betrachtet, andererseits, ist die t -Kenngrösse ein direktes Mass dafür, wie stark die Quadratsumme der Residuen zunimmt, wenn der entsprechende Regressor eliminiert wird. Dabei wird berücksichtigt, dass nach der Elimination die andern Regressionsparameter ihre Werte anpassen. Ist eine t -Kenngrösse viel kleiner als die übrigen, so kann der entsprechende Regressor eliminiert werden, ohne dass die Quadratsumme der Residuen stark zunimmt.

Die beiden Interpretationen sind insofern konsistent, als dass eine kleine t -Kenngrösse sowohl statistisch wie auch numerisch einen unwichtigen Regressor anzeigt. Allerdings kann es vorkommen, dass ein Regressor numerisch gesehen einen kleinen Beitrag liefert, obwohl er statistisch signifikant erscheint. Dieser scheinbare Widerspruch, ist meistens auf das Problem der Autokorrelation der Fehler zurückzuführen. Bei Zeitreihen sind die Fehler typischerweise nicht voneinander unabhängig, sondern unmittelbar aufeinander folgende Fehler sind korreliert. In diesem Fall sind die Annahmen der klassischen Methode der kleinsten Quadrate nicht mehr erfüllt. Die t -Kenngrössen werden dann meistens überschätzt, und einige Regressoren erscheinen fälschlicherweise als statistisch signifikant. Die numerische Interpretation der t -Kenngrössen jedoch, nämlich als Mass für das Anwachsen der Residuen, wenn ein Regressor eliminiert wird, gilt auch wenn die Fehler korreliert sind. Eine Verallgemeinerung der Methode der kleinsten Quadrate, die auch bei einer Autokorrelation der Fehler funktioniert, ist im nächsten Kapitel beschrieben.

Dass die t -Testgrössen nur jeweils für den nächsten Eliminationsschritt gelten, wurde schon früher erwähnt. Wird ein Regressor eliminiert, ändern die übrigen Koeffizienten ihren Wert und passen sich neu an. Auch die t -Testgrössen ändern sich, sodass unwichtige Regressoren plötzlich wichtig werden können und wichtige unwichtig. Dieses Phänomen tritt umso stärker auf, je mehr die Regressoren miteinander korreliert sind. Wären alle Regressoren orthogonal, könnten sie unabhängig voneinander eliminiert werden. Die gegenseitige Abhängigkeit der Regressoren ist der Grund, dass man bei der Elimination immer schrittweise vorgehen sollte, d.h. es sollte immer nur ein einzelner Regressor auf Mal eliminiert werden. Sonst können Regressoren, die am Schluss wichtig sind, schon ganz am Anfang aus dem Rennen fallen. Auch so ist nicht garantiert, dass man zur optimalen Lösung kommt. Diese kann nur gefunden werden, wenn alle möglichen Kombinationen (Untermengen) von Regressoren untersucht werden. Es gibt zwar effiziente Algorithmen, die das tun (Furnival und Wilson 1974), aber der Aufwand ist trotzdem beträchtlich, gibt es doch bei k Regressoren 2^k Kombinationen ($2^{10} = 1024, 2^{20} = 10^6$). Auch wenn nicht garantiert ist, dass man zum optimalen Resultat kommt, ist die schrittweise Elimination doch von grosser praktischer Bedeutung.

3.6.2 Einfügen

Anstatt die Regressoren schrittweise zu eliminieren, kann man sie auch schrittweise einfügen. Man fügt dem Modell schrittweise denjenigen Regressor dazu, der die Quadratsumme der Residuen am meisten reduziert. Für jeden Schritt kann die t -Testgrösse berechnet werden, die der Regressor nach hinzufügen im neuen Modell erhalten würde.

Wie bei der Elimination muss das neue Modell nicht berechnet werden, sondern die t -Testgrösse kann aus dem alten Modell bestimmt werden, wie weiter oben gezeigt wurde. Bei der Auswahl mit diesem Verfahren ergeben sich grundsätzlich die gleichen Probleme wie bei der schrittweisen Elimination, nämlich dass Regressoren, die eingeführt wurden, weil sie wichtig sind, beim Einfügen von weiteren Regressoren plötzlich unwichtig werden können. Am besten verwendet man eine kombinierte Strategie, indem man z.B. mit einem grossen Modell beginnt und schrittweise Regressoren eliminiert. In das reduzierte Modell kann man dann wieder einzelne Regressoren einfügen und das Modell wieder reduzieren. So bekommen Regressoren, die frühzeitig eliminiert wurden, eine zweite Chance. In einem Computerprogramm kann man für alle aktiven Regressoren die t -Testgrösse angeben, für die inaktiven Regressoren die t -Testgrösse, die sie nach dem Einfügen bekommen würden. Sortiert man alle Regressoren, die aktiven und die inaktiven, nach diesen t -Testgrössen, sieht man bei einer schrittweisen Elimination sofort, wenn ein Regressor, der bereits eliminiert wurde, wieder eingefügt werden sollte. Allerdings konvergiert diese Strategie nicht immer, da der wieder eingefügte Regressor die andern Regressoren beeinflusst.

4 Weitergehende Verfahren

In diesem Kapitel werden Verfahren behandelt, die angewendet werden können, wenn die gewöhnliche Verfahren versagen, entweder bei einer starken Multikollinearität oder bei einer Autokorrelation der Fehler. Bei der Multikollinearität kommen die Principal-Component-Regression und die Ridge-Regression zur Anwendung, bei der Autokorrelation der Fehler führt die verallgemeinerte Methode der kleinsten Quadrate weiter.

4.1 Standardisierte Variablen

Oft wird die Regression mit standardisierten Variablen durchgeführt. Für die bisherigen Betrachtungen war diese Form nicht nötig, aber für die folgende Principal-Component-Regression und die Ridge-Regression ist sie die richtige Form. Die standardisierten Variablen sind definiert als

$$\tilde{x}_{ij} = \frac{x_{ij} - \bar{x}_j}{\sqrt{S_{jj}}} \quad \text{mit} \quad S_{jj} = \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2, \quad (4.1)$$

und

$$\tilde{y}_i = \frac{y_i - \bar{y}}{\sqrt{S_{yy}}} \quad \text{mit} \quad S_{yy} = \sum_{i=1}^n (y_i - \bar{y})^2 = SS_{\text{Tot}}. \quad (4.2)$$

Die Umformung umfasst sowohl eine Zentrierung wie auch eine Skalierung, sodass die Vektoren Mittelwert Null und Länge Eins haben. Mit den Matrizen

$$\tilde{\mathbf{X}} = \begin{bmatrix} \tilde{x}_{11} & \cdots & \tilde{x}_{1k} \\ \tilde{x}_{21} & & \tilde{x}_{2k} \\ \vdots & & \vdots \\ \tilde{x}_{n1} & \cdots & \tilde{x}_{nk} \end{bmatrix}, \quad \tilde{\mathbf{y}} = \begin{bmatrix} \tilde{y}_1 \\ \tilde{y}_2 \\ \vdots \\ \tilde{y}_n \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}, \quad \tilde{\boldsymbol{\varepsilon}} = \begin{bmatrix} \tilde{\varepsilon}_1 \\ \tilde{\varepsilon}_2 \\ \vdots \\ \tilde{\varepsilon}_n \end{bmatrix}, \quad (4.3)$$

kann die Regression geschrieben werden als

$$\tilde{\mathbf{y}} = \tilde{\mathbf{X}}\mathbf{b} + \tilde{\boldsymbol{\varepsilon}}. \quad (4.4)$$

Da die Regressionsfunktion (Hyperebene) wie in der einfachen Regression durch den Schwerpunkt geht, verschwindet mit der Zentrierung der konstante Term. Die Matrix $\tilde{\mathbf{X}}$ enthält daher nur noch die Regressoren ohne die Einer für die Konstante und hat damit auch eine Spalte weniger als die ursprüngliche Matrix \mathbf{X} (siehe Anhang). Die Normalgleichungen sind dann

$$\tilde{\mathbf{X}}^T(\tilde{\mathbf{y}} - \tilde{\mathbf{X}}\mathbf{b}) = \mathbf{0}, \quad (4.5)$$

und die standardisierten Regressionskoeffizienten werden

$$\hat{\mathbf{b}} = (\tilde{\mathbf{X}}^T \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}^T \tilde{\mathbf{y}}. \quad (4.6)$$

Die geschätzten $\hat{\mathbf{y}}$ sind

$$\hat{\mathbf{y}} = \bar{y}\mathbf{1} + \sqrt{S_{yy}} \tilde{\mathbf{X}} \hat{\mathbf{b}}, \quad (4.7)$$

wobei $\mathbf{1}$ einen Vektor mit Einsen bezeichnet. Die Umformung von standardisierten Regressionskoeffizienten zu gewöhnlichen Regressionskoeffizienten erfolgt über die Skalierung

$$\hat{\beta}_j = \hat{b}_j \left(\frac{S_{yy}}{S_{jj}} \right)^{1/2}, \quad j = 1, 2, \dots, k. \quad (4.8)$$

Der Koeffizient für den konstanten Faktor wird

$$\hat{\beta}_0 = \bar{y} - \sum_{j=1}^k \hat{\beta}_j \bar{x}_j. \quad (4.9)$$

Die Varianz der standardisierten Regressionskoeffizienten ist

$$\text{var}(\hat{\mathbf{b}}) = \sigma^2 (\tilde{\mathbf{X}}^T \tilde{\mathbf{X}})^{-1} \quad (4.10)$$

Im Folgenden wird der Diagonalterm (j, j) der inversen Matrix $(\tilde{\mathbf{X}}^T \tilde{\mathbf{X}})^{-1}$ mit \tilde{C}_{jj} bezeichnet und es wird die im Anhang hergeleitete Beziehung $\tilde{C}_{jj} = S_{jj} C_{jj}$ benutzt. Der VIF kann mit den standardisierten Grössen wie folgt geschrieben werden:

$$\text{VIF}_j = C_{jj} S_{jj} = \tilde{C}_{jj}. \quad (4.11)$$

Die t -Testgrösse kann ebenfalls in standardisierten Variablen ausgedrückt werden:

$$t_0 = \frac{\hat{\beta}_j}{\sqrt{MS_{\text{Res}} C_{jj}}} = \frac{\hat{b}_j \sqrt{S_{yy} / S_{jj}}}{\sqrt{MS_{\text{Res}} \tilde{C}_{jj} / S_{jj}}} = \frac{\hat{b}_j \sqrt{S_{yy}}}{\sqrt{MS_{\text{Res}} \tilde{C}_{jj}}}. \quad (4.12)$$

Aus der letzten Beziehung wird auch ersichtlich, dass

$$t_0 = \frac{\hat{\beta}_j}{\text{se}(\hat{\beta}_j)} = \frac{\hat{b}_j \sqrt{S_{yy}}}{\text{se}(\hat{b}_j)}. \quad (4.13)$$

Es tritt also ein zusätzlicher Skalierungsfaktor $\sqrt{S_{yy}}$ auf.

4.2 Principal-Component-Regression

Eine Methode zur Beseitigung der Multikollinearität ist die Principal-Component-Regression. Multikollinearität bedeutet ja, dass mehrere Regressoren voneinander abhängig sind. Dies bedeutet auch, dass die Matrix $\mathbf{X}^T \mathbf{X}$ annähernd linear abhängige Spalten hat und schlecht konditioniert ist. Da die Principal-Component-Regression meistens mit standardisierten Variablen durchgeführt wird, muss man hier anstelle von $\mathbf{X}^T \mathbf{X}$ die Korrelationsmatrix $\tilde{\mathbf{X}}^T \tilde{\mathbf{X}}$ betrachten. Es ist nun möglich, durch eine Transformation die Korrelationsmatrix zu diagonalisieren und so das Problem zu orthogonalisieren.

4.2.1 Methode

Eine Eigenwertzerlegung von $\tilde{\mathbf{X}}^T \tilde{\mathbf{X}}$ führt zu

$$\tilde{\mathbf{X}}^T \tilde{\mathbf{X}} = \mathbf{V} \mathbf{\Lambda} \mathbf{V}^T, \quad (4.14)$$

wobei die Matrix \mathbf{V} die Eigenvektoren enthält und die Diagonalmatrix $\mathbf{\Lambda}$ die Eigenwerte λ_i . Die Eigenvektoren sind orthonormal, d.h. $\mathbf{V} \mathbf{V}^T = \mathbf{V}^T \mathbf{V} = \mathbf{I}$. Durch Einfügen der Einheitsmatrix kann man das Regressionsmodell neu schreiben als

$$\tilde{\mathbf{y}} = \tilde{\mathbf{X}} \mathbf{b} + \boldsymbol{\varepsilon} = \tilde{\mathbf{X}} \mathbf{V} \mathbf{V}^T \mathbf{b} + \boldsymbol{\varepsilon} = \mathbf{Z} \boldsymbol{\alpha} + \boldsymbol{\varepsilon}, \quad (4.15)$$

mit den neuen Variablen

$$\mathbf{Z} = \tilde{\mathbf{X}} \mathbf{V} \quad (4.16)$$

und den neuen Regressionskoeffizienten

$$\boldsymbol{\alpha} = \mathbf{V}^T \mathbf{b}. \quad (4.17)$$

Die (nicht normierte) Korrelationsmatrix der transformierten Regressoren

$$\mathbf{Z}^T \mathbf{Z} = \mathbf{V}^T \tilde{\mathbf{X}}^T \tilde{\mathbf{X}} \mathbf{V} = \mathbf{\Lambda} \quad (4.18)$$

ist nun eine Diagonalmatrix, d.h. die transformierten Variablen \mathbf{z}_i sind orthogonal. Die Koeffizienten für die transformierten Regressoren \mathbf{Z} werden

$$\begin{aligned} \hat{\boldsymbol{\alpha}} &= (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \tilde{\mathbf{y}} \\ &= \mathbf{\Lambda}^{-1} \mathbf{Z}^T \tilde{\mathbf{y}}. \end{aligned} \quad (4.19)$$

Die geschätzten $\hat{\mathbf{y}}$ sind dann

$$\hat{\mathbf{y}} = \bar{y} \mathbf{1} + \sqrt{S_{yy}} \mathbf{Z} \hat{\boldsymbol{\alpha}} = \bar{y} \mathbf{1} + \sqrt{S_{yy}} \mathbf{Z} \mathbf{\Lambda}^{-1} \mathbf{Z}^T \tilde{\mathbf{y}}. \quad (4.20)$$

Die obigen Formeln werden etwas klarer, wenn man einzelne Eigenvektoren betrachtet. Ein Eigenvektor kann geschrieben werden als

$$\mathbf{v}_i = \begin{bmatrix} v_{1i} \\ v_{2i} \\ \vdots \\ v_{ni} \end{bmatrix} \quad (4.21)$$

Die Spalten von \mathbf{Z} sind die transformierten Variablen \mathbf{z}_i und werden auch Principal Components genannt. Sie sind Linearkombinationen der ursprünglichen Regressoren \mathbf{x}_j , nämlich [aus (4.16)]

$$\mathbf{z}_i = \tilde{\mathbf{X}} \mathbf{v}_i = v_{1i} \tilde{\mathbf{x}}_1 + v_{2i} \tilde{\mathbf{x}}_2 + \cdots + v_{ki} \tilde{\mathbf{x}}_k. \quad (4.22)$$

Da $\mathbf{\Lambda}$ diagonal ist, sind die Gleichungen für die einzelnen Regressionskoeffizienten entkoppelt [aus (4.19)]

$$\hat{\alpha}_i = \frac{\mathbf{z}_i^T \tilde{\mathbf{y}}}{\lambda_i}. \quad (4.23)$$

Die neuen Regressionskoeffizienten können durch die alten ausgedrückt werden als [aus (4.17)]

$$\alpha_i = \mathbf{v}_i^T \mathbf{b} = v_{1i} b_1 + v_{2i} b_2 + \cdots v_{ki} b_k. \quad (4.24)$$

Aus der linearen Algebra ist bekannt, dass eine Matrix schlecht konditioniert ist, wenn der kleinste Eigenwert gegenüber dem grössten Eigenwert sehr klein ist. Die Konditionszahl der Matrix $\mathbf{X}^T \mathbf{X}$ ist als $\kappa = \lambda_{\max} / \lambda_{\min}$ definiert, wobei λ_{\max} und λ_{\min} der grösste und der kleinste Eigenwert von $\mathbf{X}^T \mathbf{X}$ sind. Eine grosse Konditionszahl bedeutet eine schlecht konditionierte Matrix. Um die Multikollinearität aufzuheben, werden nun in der Principal-Component-Regression Eigenvektoren mit kleinen Eigenwerten weggelassen. Im Allgemeinen nimmt die Quadratsumme der Residuen dadurch nicht wesentlich zu. Der Zuwachs hängt nicht nur von den Regressoren ab sondern auch von $\tilde{\mathbf{y}}$. Eine ausführlichere Erklärung bietet der Abschnitt im Anhang über die Singulärwertzerlegung. Eine statistische Interpretation folgt in Abschnitt 4.2.3.

Für eine einheitliche Implementierung in einem Computerprogramm kann die folgende Überlegung nützlich sein. Das Weglassen von einigen Eigenvektoren kann durch Weglassen der entsprechenden Spalten in \mathbf{V} und Diagonaltermen in $\mathbf{\Lambda}$ erfolgen. Bezeichnet man die reduzierten Matrizen mit \mathbf{V}_1 und $\mathbf{\Lambda}_1$, sind die transformierten Regressionskoeffizienten

$$\hat{\boldsymbol{\alpha}} = \mathbf{\Lambda}_1^{-1} \mathbf{V}_1^T \tilde{\mathbf{X}}^T \tilde{\mathbf{y}}. \quad (4.25)$$

Die geschätzten $\hat{\mathbf{y}}$ sind dann

$$\begin{aligned} \hat{\mathbf{y}} &= \bar{y} \mathbf{1} + \sqrt{S_{yy}} (\tilde{\mathbf{X}} \mathbf{V}_1) (\mathbf{\Lambda}_1^{-1} \mathbf{V}_1^T \tilde{\mathbf{X}}^T \tilde{\mathbf{y}}) \\ &= \bar{y} \mathbf{1} + \sqrt{S_{yy}} \tilde{\mathbf{X}} (\mathbf{V}_1 \mathbf{\Lambda}_1^{-1} \mathbf{V}_1^T) \tilde{\mathbf{X}}^T \tilde{\mathbf{y}}. \end{aligned} \quad (4.26)$$

Gegenüber der ursprünglichen Formulierung wird also einfach die Matrix $(\tilde{\mathbf{X}}^T \tilde{\mathbf{X}})^{-1}$ durch $\mathbf{V}_1 \mathbf{\Lambda}_1^{-1} \mathbf{V}_1^T$ ersetzt.

4.2.2 Beispiel

Ein einfaches Beispiel soll zeigen, wie die Principal-Component-Regression funktioniert. Gegeben seien die Matrizen

$$\mathbf{y} = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \quad \text{und} \quad \mathbf{X} = \begin{bmatrix} 1 & 0.9 \\ 0 & 0.2 \\ -1 & -1.1 \end{bmatrix}. \quad (4.27)$$

Um eine möglichst einfache Darstellung zu haben sind die Variablen zentriert und die Konstante fällt damit weg. Die Variablen sind nicht normiert, was aber nichts an der Aussage des Beispiels ändert. Die erste Spalte von \mathbf{X} ist gerade gleich wie \mathbf{y} , die zweite Spalte ist nur leicht verschieden. Wenn man die VIF-Werte berechnet, findet man, dass beide um 35 sind, dass man also eine Kollinearität hat. Es ist klar, dass $\boldsymbol{\beta}^T = [1 \ 0]$ die exakte Lösung ist, und dass damit die Residuen verschwinden. Die zweite Spalte von \mathbf{X} stört eigentlich nur und verursacht die Kollinearität. Umgekehrt, wäre \mathbf{y} gleich der

zweiten Spalte, so würde die Berechnung jene auswählen. Es hängt also nur von Kleinigkeiten im \mathbf{y} -Vektor ab, welcher der beiden Regressoren wie wichtig ist.

Die Eigenwertzerlegung liefert die Eigenwerte $\lambda_1 = 4.03$ und $\lambda_2 = 0.03$. Man kann also den zweiten Eigenvektor weglassen und nur den ersten, $\mathbf{v}_1^T = [0.702 \quad 0.712]$ verwenden. Da nur ein Eigenvektor gewählt wird, müssen die Regressionskoeffizienten proportional zu diesem sein. Die Berechnung liefert $\boldsymbol{\beta}^T = [0.493 \quad 0.5]$, und die Quadratsumme der Residuen wird $SS_{\text{Res}} = 0.015$. Eine kleine Änderung des \mathbf{y} -Vektors würde an diesem Resultat wenig ändern, da die Koeffizienten immer proportional zum Eigenvektor bleiben müssen.

Das Beispiel zeigt auch, dass es nicht sinnvoll ist, viele ähnliche Regressoren ins Modell einzuführen und zu hoffen, dass die Berechnung die besten auswählt. Die Principal-Component-Regression kann das nicht, sie findet eher einen Kompromiss bei dem der Einfluss von allen ähnlichen Regressoren berücksichtigt wird. Man muss sich auch immer bewusst sein, dass die Eigenwertzerlegung ohne den \mathbf{y} -Vektor durchgeführt wird und daher die Regressoren unabhängig von \mathbf{y} transformiert werden.

Etwas anders verhält sich das Beispiel wenn man

$$\mathbf{y} = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \quad (4.28)$$

setzt. Die beiden Regressoren zeigen eigentlich in eine ganz anderer Richtung und eine Approximation kann nur aus ihrer Differenz konstruiert werden. Die klassische Methode berechnet $\boldsymbol{\beta}^T = [5.5 \quad -5]$ und die Residuen sind wieder Null. Die Principal-Component-Regression liefert $\boldsymbol{\beta}^T = [0.209 \quad 0.212]$ und $SS_{\text{Res}} = 1.642$. Die Regressionskoeffizienten müssen immer noch proportional zum Eigenvektor sein und können daher nicht verschiedene Vorzeichen annehmen wie bei der klassischen Methode. Damit kann aber auch der \mathbf{y} -Vektor nicht mehr gut approximiert werden, was in diesem Fall sinnvoll ist, da man gerade verhindern will, dass Regressoren ohne Aussagekraft durch eine Kombination zu viel Einfluss bekommen.

4.2.3 Interpretation

Zuerst soll hier noch etwas deutlicher ausgeführt werden, wie die Eigenwerte und die Multikollinearität zusammenhängen. Zum einen kann man die VIF-Werte durch die Eigenwerte ausdrücken. Mit $\tilde{\mathbf{C}} = (\tilde{\mathbf{X}}^T \tilde{\mathbf{X}})^{-1} = \mathbf{V} \boldsymbol{\Lambda}^{-1} \mathbf{V}^T$ folgt

$$\text{VIF}_j = \tilde{C}_{jj} = \sum_{i=1}^k \frac{v_{ij}^2}{\lambda_i}. \quad (4.29)$$

Kleine Eigenwerte führen tatsächlich zu grossen VIF-Werten. Eine analoge Überlegung kann für die Varianz der Regressionskoeffizienten gemacht werden:

$$\text{var}(\hat{b}_j) = \sigma^2 \tilde{C}_{jj} = \sigma^2 \sum_{i=1}^k \frac{v_{ij}^2}{\lambda_i}. \quad (4.30)$$

Kleine Eigenwerte sind also auch für eine grosse Varianz der Regressionskoeffizienten verantwortlich. Durch das Weglassen kleiner Eigenwerte werden die Regressionskoeffizienten verändert und sind damit verzerrte (biased) Schätzwerte. Dafür wird die Varianz verkleinert, was insbesondere für die Prognose von Vorteil sein kann. Man beachte, dass es nach dem Satz von Gauss-Markov nicht möglich ist die Varianz der Regressionskoeffizienten zu verkleinern, ohne dass die Schätzung verzerrt wird.

Man kann aber auch die Varianz der transformierten Regressionskoeffizienten betrachten. Diese ist durch folgenden Ausdruck gegeben:

$$\text{var}(\hat{\alpha}) = \sigma^2 (\mathbf{Z}^T \mathbf{Z})^{-1} = \sigma^2 \mathbf{\Lambda}^{-1}. \quad (4.31)$$

Für einen einzelnen Koeffizienten heisst das

$$\text{var}(\hat{\alpha}_i) = \frac{\sigma^2}{\lambda_i}. \quad (4.32)$$

Wenn ein Eigenwert sehr klein ist, ist diese Varianz sehr gross, was bedeutet, dass die entsprechende Variable \mathbf{z}_i fast eine Konstante ist. Das heisst

$$\mathbf{z}_i = v_{1i} \tilde{\mathbf{x}}_1 + v_{2i} \tilde{\mathbf{x}}_2 + \dots + v_{ki} \tilde{\mathbf{x}}_k \approx \text{konst}, \quad (4.33)$$

was nichts anderes ist als die Definition der Multikollinearität. Aus der Gleichung können direkt die Beziehungen, die die Regressoren verbinden, herausgelesen werden. Wenn aber \mathbf{z}_i fast eine Konstante ist, dann kann man sie auch einfach weglassen. Das ist genau das, was die Principal-Component-Regression tut.

Das Weglassen eines Eigenvektors kann als Einführen einer zusätzlichen Nebenbedingung interpretiert werden (Chatterjee und Price 1977, Kapitel 8.4). Weglassen des Eigenvektors i bedeutet dessen Koeffizienten zu Null setzen:

$$\alpha_i = v_{1i} b_1 + \dots + v_{ki} b_k = 0. \quad (4.34)$$

Daraus erhält man eine Bedingung für die Regressionskoeffizienten b_j . Oft sind nur wenige der Faktoren massgebend. Ist z.B. $\alpha_i \approx v_{1i} b_1 + v_{2i} b_2 \approx 0$, kann man diese Bedingung in die Regressionsgleichung einsetzen und erhält nebst anderen Termen

$$b_1 \tilde{\mathbf{x}}_1 + b_2 \tilde{\mathbf{x}}_2 = b_1 \tilde{\mathbf{x}}_1 - (v_{1i} / v_{2i}) b_1 \tilde{\mathbf{x}}_2 = b_1 (\tilde{\mathbf{x}}_1 - (v_{1i} / v_{2i}) \tilde{\mathbf{x}}_2). \quad (4.35)$$

Der letzte Klammerausdruck kann als neue Variable aufgefasst werden. Man kann also $\tilde{\mathbf{x}}_1$ und $\tilde{\mathbf{x}}_2$ zu einer einzigen Variablen kombinieren. Mit dem Einführen dieser kombinierten Variablen ist dann auch die entsprechende Kollinearität beseitigt. Man kann diese Technik benützen um z.B. Temperaturen, die stark korreliert sind, zusammenzufassen. Die Methode wird aber bald einmal komplizierter, wenn viele Regressoren an der Multikollinearität beteiligt sind. Es müssen dann mehrere Eigenvektoren weggelassen werden, und anstatt einer Beziehungsgleichung hat man ein Gleichungssystem.

4.2.4 t-Test für transformierte Regressoren

Grundsätzlich führt man eine Principal-Component-Regression durch, um Multikollinearität zu beseitigen. Man sollte daher nur die Eigenvektoren mit dem kleinsten Eigenwerten weglassen. Andererseits interessiert es aber auch, welche Eigenvektoren die

Residuen am wenigsten vergrössern, wenn sie weggelassen werden. Um dies festzustellen, können analog zur ursprünglichen Formulierung t -Testgrössen für die transformierten Regressionskoeffizienten berechnet werden:

$$t_0 = \frac{\hat{\alpha}_j \sqrt{S_{yy}}}{\text{se}(\hat{\alpha}_j)} = \frac{\lambda_j^{-1} \mathbf{v}_j^T \tilde{\mathbf{X}}^T \tilde{\mathbf{y}} \sqrt{S_{yy}}}{\sqrt{MS_{\text{Res}}/\lambda_j}} = \frac{\mathbf{v}_j^T \tilde{\mathbf{X}}^T \tilde{\mathbf{y}} \sqrt{S_{yy}/\lambda_j}}{\sqrt{MS_{\text{Res}}}}. \quad (4.36)$$

Im Anhang wird gezeigt, dass beim Weglassen eines Eigenvektors \mathbf{v}_i der Zuwachs in der Quadratsumme der Residuen

$$\Delta SS_{\text{Res}} = S_{yy} \frac{(\mathbf{v}_i^T \tilde{\mathbf{X}}^T \tilde{\mathbf{y}})^2}{\lambda_i} \quad (4.37)$$

beträgt. Damit kann man leicht verifizieren, dass auch die bekannte Beziehung zwischen der t -Kenngrösse und der partiellen F -Kenngrösse gilt:

$$F_0 = \frac{\Delta SS_{\text{Res}}}{MS_{\text{Res}}} = t_0^2. \quad (4.38)$$

Die t -Testgrössen für kleine Eigenwerte sind im Allgemeinen auch klein. (Dies ist aus der Formel nicht ersichtlich.) Wenn also die entsprechenden Eigenvektoren weggelassen werden, ist das im Allgemeinen nur mit einem kleinen Zuwachs der Residuen verbunden. Allerdings sind nicht immer die kleinsten t_0 mit den kleinsten Eigenwerten assoziiert. Der Wert von t_0 hängt auch von $\tilde{\mathbf{y}}$ ab, das in der Eigenwertzerlegung nicht beteiligt ist.

4.2.5 Ursprüngliche Regressionskoeffizienten und t -Test

Die Überlegungen in diesem Abschnitt stützen sich hauptsächlich auf die Publikation Mansfield et al. (1977). Die ursprünglichen Regressionskoeffizienten können aus den transformierten Koeffizienten berechnet werden als

$$\hat{\mathbf{b}}_{\text{PC}} = \mathbf{V}_1 \hat{\boldsymbol{\alpha}} = (\mathbf{V}_1 \boldsymbol{\Lambda}_1^{-1} \mathbf{V}_1^T) \tilde{\mathbf{X}}^T \tilde{\mathbf{y}}. \quad (4.39)$$

Dies ist die ursprüngliche Formel mit $(\tilde{\mathbf{X}}^T \tilde{\mathbf{X}})^{-1}$ ersetzt durch $\mathbf{V}_1 \boldsymbol{\Lambda}_1^{-1} \mathbf{V}_1^T$. Man kann für jeden Regressionskoeffizienten \hat{b}_j herausfinden, um wie viel die Quadratsumme der Residuen zunimmt, wenn der entsprechende Regressor eliminiert wird. Dabei werden die Eigenvektoren nicht neu berechnet sondern einfach der Regressor $\tilde{\mathbf{x}}_j$ von den Linearkombinationen weggelassen:

$$\mathbf{z}_i = v_{1i} \tilde{\mathbf{x}}_1 + v_{2i} \tilde{\mathbf{x}}_2 + \cdots + v_{j-1,i} \tilde{\mathbf{x}}_{j-1} + v_{j+1,i} \tilde{\mathbf{x}}_{j+1} + \cdots + v_{ki} \tilde{\mathbf{x}}_k. \quad (4.40)$$

Im Anhang wird gezeigt, dass der Beitrag der Regressors $\tilde{\mathbf{x}}_j$ zur Quadratsumme der Residuen

$$\Delta SS_{\text{Res}} = \frac{\hat{b}_j^2 S_{yy}}{(\mathbf{V}_1 \boldsymbol{\Lambda}_1^{-1} \mathbf{V}_1^T)_{jj}} \quad (4.41)$$

beträgt. Analog zur ursprünglichen Formulierung kann man daraus eine (Pseudo-) t -Testgrösse herleiten, die zwar noch den Zuwachs der Residuen beschreibt, aber nicht

mehr die statistische Bedeutung der t -Kenngrösse besitzt. Aus $t_0^2 = F_0 = \Delta SS_{\text{Res}} / MS_{\text{Res}}$ findet man

$$t_0 = \frac{\hat{b}_j \sqrt{S_{yy}}}{\sqrt{MS_{\text{Res}} (\mathbf{V}_1 \mathbf{\Lambda}_1^{-1} \mathbf{V}_1^T)_{jj}}}. \quad (4.42)$$

Dies ist der gleiche Ausdruck wie bei der normalen Regression, wenn man $(\tilde{\mathbf{X}}^T \tilde{\mathbf{X}})^{-1}$ durch $\mathbf{V}_1 \mathbf{\Lambda}_1^{-1} \mathbf{V}_1^T$ ersetzt.

Man sieht also, dass bei der Principal-Component-Regression die gleichen Formeln benutzt werden können, wie bei der normalen Regression, wenn man $(\tilde{\mathbf{X}}^T \tilde{\mathbf{X}})^{-1}$ durch $\mathbf{V}_1 \mathbf{\Lambda}_1^{-1} \mathbf{V}_1^T$ ersetzt. Dies ist vor allem auch bei der Implementierung ein Vorteil, da man den gleichen Code für beide Methoden verwenden kann.

4.2.6 Diskussion

Die Principal-Component-Regression eignet sich gut für die Beseitigung einer Multikollinearität. Die Voraussetzung ist allerdings, dass nicht allzu viele Multikollinearitäten beseitigt werden müssen. Sind nur wenige Regressoren voneinander abhängig, so findet man typischerweise einige wenige Eigenwerte, die viel kleiner sind als die übrigen. Wenn man sie weglässt, vergrößert sich die Quadratsumme der Residuen nur unwesentlich. Ein anderes Bild zeigt sich meistens, wenn sehr viele Multikollinearitäten vorhanden sind. Die Eigenwerte sind dann typischerweise viel gleichmässiger verteilt, und es ist nicht klar, wo man abschneiden soll. Man muss auf jeden Fall viele Eigenvektoren weglassen, was zu einem übermässigen Anstieg der Quadratsumme der Residuen führen kann. Das bedeutet aber auch, dass die verbleibenden Eigenvektoren nicht mehr in der Lage sind, die unabhängige Variable gut zu beschreiben.

Bei der Variablenauswahl mit Multikollinearität ergeben sich mehrere Möglichkeiten, da sowohl die Eigenvektoren wie auch die Regressoren ausgewählt werden können. Zwei Varianten sollen hier etwas ausführlicher diskutiert werden.

Man kann sich auf den Standpunkt stellen, dass man die Variablentransformation aus der Principal-Component-Regression beibehält und nur noch die neuen z -Variablen betrachtet. Man kann dann zuerst jene mit den kleinsten Eigenwerten weglassen, um die Multikollinearitäten zu beseitigen. Anschliessend kann man die z -Variablen mit den kleinsten t -Testgrössen weglassen, um das Modell zu verkleinern. Da alle z -Variablen zueinander orthogonal sind, muss auch nicht schrittweise vorgegangen werden. So gut das klingt, das Problem ist damit nicht gelöst. Das Problem beim Weglassen der z -Variablen mit den kleinsten Eigenwerten ist wieder, dass die verbleibenden Eigenvektoren nicht mehr geeignete Regressoren sind und daher grosse Residuen entstehen. Auch das Weglassen von z -Variablen mit einer kleinen t -Testgrösse ist nicht so vorteilhaft, wie man das auf den ersten Blick annehmen könnte. Die Residuen werden zwar nur wenig vergrößert, aber die ursprünglichen Regressoren sind immer noch alle im Modell. Macht man die Rücktransformation, so sieht man nichts mehr davon, dass die z -Variablen mit kleinen t -Testgrössen weggelassen wurden, ausser dass die Residuen leicht zugenommen haben.

Eine andere Strategie ist die, dass man zuerst mit der Principal-Component Regression die Multikollinearität beseitigt. Dann sind alle ursprünglichen Regressoren mehr oder weniger voneinander unabhängig und sie können nach ihren t -Testgrössen ausgewählt werden. Diese Strategie wird in Mansfield et al. (1977) vorgeschlagen. Dort wird auch empfohlen, nach der Elimination die Eigenvektoren wieder neu zu berechnen. Die in dieser Studie betrachteten Beispiele haben aber gezeigt, dass diese Strategie oft ungünstig ist. Sind am Anfang viele Regressoren an der Multikollinearität beteiligt, müssen viele Eigenvektoren weggelassen werden, was einen markanten Anstieg der Quadratsumme der Residuen zur Folge hat. Werden dann die Regressoren eliminiert, steigt der Fehler weiter an. Eine neue Berechnung der Eigenvektoren kann zwar den Fehler wieder verkleinern, aber trotzdem scheint diese Strategie oft die falschen Regressoren zu eliminieren.

Trotz vielen Überlegungen und Berechnungen konnte keine allgemein anwendbare Strategie gefunden werden, wie die Principal-Component-Regression für die Variablenauswahl eingesetzt werden kann. Es scheint, dass es bei sehr vielen, stark voneinander abhängigen Regressoren gar nicht mehr möglich ist, die Eigenvektoren so auszuwählen, dass die Multikollinearität beseitigt wird und die Residuen trotzdem klein bleiben. Etwas bessere Möglichkeiten bei der Variablenauswahl bietet die Ridge-Regression.

4.3 Ridge-Regression

Eine andere, einfachere Möglichkeit zur Beseitigung der Multikollinearität ist die so genannte Ridge-Regression. Die fast singuläre Matrix $(\tilde{\mathbf{X}}^T \tilde{\mathbf{X}})$ wird leicht modifiziert indem ein kleiner Diagonalterm hinzugefügt wird, nämlich $(\tilde{\mathbf{X}}^T \tilde{\mathbf{X}}) + k\mathbf{I}$. Die Regressionskoeffizienten werden dadurch:

$$\hat{\mathbf{b}}_R = (\tilde{\mathbf{X}}^T \tilde{\mathbf{X}} + k\mathbf{I})^{-1} \tilde{\mathbf{X}}^T \tilde{\mathbf{y}}. \quad (4.43)$$

Wie bei der Principal-Component-Regression ist dies ein verzerrter (biased) Schätzwert für die Regressionskoeffizienten. Der Parameter k heisst denn auch Verzerrungsparameter (biasing parameter). Der Nachteil der Verzerrung wird durch die kleinere Varianz der Regressionskoeffizienten aufgewogen (Montgomery, S. 349).

Der grosse Vorteil dieser Methode ist die einfache Implementierung. Man kann nämlich einfach die Daten wie folgt erweitern:

$$\mathbf{X}_A = \begin{bmatrix} \tilde{\mathbf{X}} \\ \sqrt{k}\mathbf{I} \end{bmatrix} \quad \text{und} \quad \mathbf{y}_A = \begin{bmatrix} \tilde{\mathbf{y}} \\ \mathbf{0} \end{bmatrix}. \quad (4.44)$$

Mit diesen Daten lässt sich dann die Standard-Regression durchführen. Der Nachteil, der Methode ist der, dass der Parameter k vom Benutzer gewählt werden muss. Es gibt zwar verschiedene Regeln, die aber nicht generell akzeptiert sind. Der Wert sollte möglichst klein sein, da sonst die Residuen zu stark ansteigen, andererseits muss er genügend gross sein um die Multikollinearität zu beseitigen. Eine Möglichkeit ist, verschiedene Werte im Bereich von 0 bis 1 auszuprobieren und zu sehen, ob und wo sich die Regressionskoeffizienten stabilisieren (graphisch als so genannte Ridge Trace). In Montgomery (S. 362) werden auch Regeln angegeben, wie aus dem Verhalten der

Regressionskoeffizienten bei wachsendem Verzerrungsparameter Hinweise für die Variablenauswahl gewonnen werden können.

4.3.1 Diskussion

In den Beispielen wurde festgestellt, dass der Verzerrungsparameter immer relativ gross gewählt werden musste um die VIF-Werte unter 10 zu bringen. Der dadurch verursachte Anstieg in der Quadratsumme der Residuen war meist inakzeptabel und viel grösser als bei der Principal-Component-Regression. Es hat sich aber gezeigt, dass ein kleiner Verzerrungsparameter (z.B. $k = 0.005$) die Multikollinearität soweit reduziert (VIF-Werte um 100), dass dadurch die schrittweise Elimination der Regressoren stark verbessert wird. Allerdings können verschiedene Verzerrungsparameter zu verschiedenen Modellen führen, sodass es sich empfiehlt, mehrere Varianten zu untersuchen. Oft ist im reduzierten Modell keine Multikollinearität mehr vorhanden und wenn noch eine da ist, kann sie mit der Principal-Component-Regression aufgehoben werden.

4.4 Autokorrelation der Fehler

Zeigt der Durbin-Watson-Test oder das Korrelogramm eine Autokorrelation der Fehler an, so sind die Regressionskoeffizienten zwar noch unverzerrt, aber nicht mehr effizient, d.h. sie haben nicht mehr minimale Varianz. Ebenso werden die t -Kenngrössen falsch berechnet. Eine allgemeinere Methode, die auch mit korrelierten Fehlern funktioniert, ist die verallgemeinerte Methode der kleinsten Quadrate (Generalized Least Squares). Diese ist im Anhang beschrieben. Hier wird nur der einfachste Fall der Autokorrelation erster Ordnung untersucht, der durch das folgende Fehlermodell beschrieben werden kann:

$$\varepsilon_t = \rho \varepsilon_{t-1} + u_t. \quad (4.45)$$

In jedem Zeitschritt t besteht der Fehler aus einem Anteil ρ des vorangehenden Fehlers und einem neuen Fehler u_t . Die Fehler u_t sind normal verteilt und unkorreliert. Für diesen Fall reduziert sich die Methode auf eine verallgemeinerte Differenzenbildung (generalized differencing). Die Methode wird hier für einen einzelnen Regressor beschrieben, wobei die Verallgemeinerung auf mehrere Regressoren ohne Schwierigkeiten ersichtlich ist. Das Modell

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t \quad (4.46)$$

kann so transformiert werden, dass die Annahme eines unkorrelierten Fehlers erfüllt ist. Dazu wird die neue Variable $y_t^* = y_t - \rho y_{t-1}$ gebildet. Diese ist

$$\begin{aligned} y_t^* &= y_t - \rho y_{t-1} \\ &= \beta_0 + \beta_1 x_t + \varepsilon_t - \rho(\beta_0 + \beta_1 x_{t-1} + \varepsilon_{t-1}) \\ &= \beta_0(1 - \rho) + \beta_1(x_t - \rho x_{t-1}) + \varepsilon_t - \rho \varepsilon_{t-1} \\ &= \beta_0(1 - \rho) + \beta_1 x_t^* + u_t. \end{aligned} \quad (4.47)$$

Das transformierte Problem lautet also

$$y_t^* = \beta_0(1 - \rho) + \beta_1 x_t^* + u_t \quad (4.48)$$

mit

$$y_t^* = y_t - \rho y_{t-1} \quad \text{und} \quad x_t^* = x_t - \rho x_{t-1}. \quad (4.49)$$

Der Fehler u_t erfüllt die Annahmen, sodass die gewöhnliche Methode der kleinsten Quadrate auf das transformierte Problem angewendet werden kann. Der Wert von ρ muss geschätzt werden. Dazu wird mit den Residuen e_t das Modell $e_t = \rho e_{t-1} + u_t$ aufgestellt und ρ durch eine einfache Regression bestimmt (Montgomery, S. 496):

$$\hat{\rho} = \frac{\sum_{t=2}^n e_t e_{t-1}}{\sum_{t=1}^n e_t^2}. \quad (4.50)$$

Dies ist nebenbei bemerkt gerade der erste Wert des Korrelogramms. Diese Methode zur Bestimmung von ρ wird in der Literatur auch nach Cochrane und Orcut (1949) benannt.

Bei der Differenzenbildung fallen die Werte für den ersten Zeitschritt weg, da eine frühere Beobachtung fehlt. In der Literatur wird vorgeschlagen, den Fehler ε_1 wie folgt zu wählen (Kmenta 1997, S. 302)

$$\varepsilon_1 = \frac{u_1}{\sqrt{1-\rho^2}}. \quad (4.51)$$

Die Transformation für den ersten Zeitschritt ist damit etwas anders als für die nachfolgenden Zeitschritte, nämlich

$$\begin{aligned} y_t^* &= y_t \sqrt{1-\rho^2} \\ &= \beta_0 \sqrt{1-\rho^2} + \beta_1 x_t \sqrt{1-\rho^2} + \varepsilon_t \sqrt{1-\rho^2} \\ &= \beta_0 \sqrt{1-\rho^2} + \beta_1 x_t^* + u_t. \end{aligned} \quad (4.52)$$

Das transformierte Problem für den ersten Zeitschritt ist also

$$y_1^* = \beta_0 \sqrt{1-\rho^2} + \beta_1 x_1^* + u_1 \quad (4.53)$$

mit

$$y_1^* = y_1 \sqrt{1-\rho^2} \quad \text{und} \quad x_1^* = x_1 \sqrt{1-\rho^2} \quad (4.54)$$

Wird der erste Zeitschritt ebenfalls berücksichtigt, ändert sich die Schätzung des Autokorrelationsparameters wie folgt:

$$\hat{\rho} = \frac{\sum_{t=2}^n e_t e_{t-1}}{\sum_{t=3}^n e_{t-1}^2} \quad (4.55)$$

Diese Methode wird nach Prais-Winston benannt (Kmenta 1997, S. 318). Im Anhang ist eine noch etwas kompliziertere Berechnung nach Pesaran und Slater (1980) angegeben, die auf einer Maximum-Likelihood-Formulierung basiert.

Mit dem geschätzten $\hat{\rho}$ werden durch verallgemeinerte Differenzenbildung die transformierten Variablen gebildet. Da auch β_0 mit einem Faktor versehen ist, muss

dieser Faktor ebenfalls berücksichtigt werden. Die Regression für das transformierte Problem ist

$$y_t^* = \beta_0 w_t^* + \beta_1 x_t^* + u_t \quad (4.56)$$

Für den ersten Zeitschritt ist

$$y_1^* = \sqrt{1 - \hat{\rho}^2} y_1 \quad \text{und} \quad x_1^* = \sqrt{1 - \hat{\rho}^2} x_1 \quad \text{und} \quad w_1^* = \sqrt{1 - \hat{\rho}^2} \quad (4.57)$$

und für alle andern Zeitschritte ist

$$y_t^* = y_t - \hat{\rho} y_{t-1} \quad \text{und} \quad x_t^* = x_t - \hat{\rho} x_{t-1} \quad \text{und} \quad w_t^* = 1 - \hat{\rho}. \quad (4.58)$$

Mit den neuen Variablen werden dann mit der gewöhnlichen Methode der kleinsten Quadrate die Regressionsparameter β_0 und β_1 bestimmt. Dabei muss in der Matrix \mathbf{X} die erste Spalte nicht mit Einsen sondern mit w_t^* aufgefüllt werden. Diese Werte sind alle gleich ausser dem ersten. Mit diesem Vorgehen werden nicht nur direkt die Regressionsparameter β_0 und β_1 sondern auch die t -Kenngrössen richtig berechnet.

Mit den so berechneten Regressionsparametern kann die Regressionsgerade in den ursprünglichen Variablen geschrieben werden. Mit den Residuen aus dieser Regression kann ein neuer Wert für $\hat{\rho}$ berechnet werden. Nach ein paar Iterationen sollte dieser Wert konvergieren.

Ein Spezialfall ist $\hat{\rho} = 1$. Obwohl damit einige der obigen Formeln ungültig werden, kann dieser Fall noch behandelt werden und wird auch in der Literatur beschrieben. Im transformierten Problem wird die Konstante dann Null,

$$y_t^* = \beta_1 x_t^* + u_t \quad (4.59)$$

und man muss die Regression ohne Konstante durchführen. Für das ursprüngliche Problem wird die Konstante dann so bestimmt, dass die Regressionsgerade durch den Schwerpunkt geht:

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}. \quad (4.60)$$

Die neuen Regressionskoeffizienten sind natürlich leicht anders als jene, die direkt berechnet werden. Die Quadratsumme der Residuen (am ursprünglichen Problem betrachtet) ist nicht mehr minimal sondern leicht grösser. Dafür ist die Varianz minimal, was sich positiv auf die Prognose auswirken sollte. In manchen Beispielen wurde beobachtet, dass die Residuen in der Regressionsperiode etwas grösser, in der Prognoseperiode dafür etwas kleiner waren als jene, die mit der gewöhnlichen Methode berechnet wurden. Johnston (1963, S. 191) gibt für die Verbesserung der Varianz folgende ungefähre Beziehung an:

$$\frac{\text{var}(\hat{\beta}')}{\text{var}(\hat{\beta})} = \frac{1 - \rho^2}{1 + \rho^2} \quad (4.61)$$

wobei $\hat{\beta}'$ den Regressionskoeffizienten nach der verallgemeinerten Methode und $\hat{\beta}$ jenen nach der gewöhnlichen Methode bezeichnet. Für $\rho = 0.5$ wird also mit der verallgemeinerten Methode die Varianz gegenüber der klassischen Methode auf 60% redu-

ziert, bei $\rho = 0.9$ sogar auf 10%. Eine so grosse Verbesserung kann natürlich nur erreicht werden, wenn die Autokorrelation der Fehler tatsächlich erster Ordnung ist, und wenn ρ richtig geschätzt wird. In den Beispielen war die Verbesserung viel weniger deutlich oder konnte überhaupt nicht festgestellt werden.

4.4.1 Diskussion

Insgesamt ist diese verallgemeinerte Methode bei einer Autokorrelation sehr viel versprechend, aber sie zeigt in ihrer einfachsten Form noch einige ungelöste Probleme. Bei einigen Beispielen konvergierte die iterative Methode zur Bestimmung von ρ nicht oder die Lösung war unbrauchbar, weil die Residuen viel zu gross waren. Das kann mehrere Gründe haben. Einmal können die Fehler auch eine Korrelation höherer Ordnung aufweisen. Zum andern handelt es sich um eine nichtlineare Regression und es ist nicht garantiert, dass die einfache Methode, die verwendet wurde, auch konvergiert. Aufwändigere Verfahren gehen von einer Maximum-Likelihood-Formulierung aus, bei der die Regressionskoeffizienten und ρ gleichzeitig berechnet werden. In den Beispielen konnte in ein paar Problemfällen eine Optimierung von Hand durchgeführt werden. Für eine allgemeine Anwendung müsste jedoch der Algorithmus noch verbessert werden.

Sowohl die Durbin-Watson-Kenngrösse und das Korrelogramm, wie auch die Transformation zur Beseitigung der Autokorrelation der Fehler gehen von Beobachtungen mit regelmässigen Zeitschritten aus. Kleinere Abweichungen sollten kein Problem sein, aber bei ganz unregelmässigen Zeitschritten, können die Methoden wohl kaum mehr angewendet werden.

Die Methode ist ähnlich wie, aber verschieden von einem autoregressiven Ansatz $y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 x + \varepsilon$. Der autoregressive Ansatz ist theoretisch schwieriger zu behandeln, da nun auch die unabhängige Variable y_{t-1} eine Zufallsvariable ist. Bei der vorgeschlagenen Methode werden zwar auch Differenzen der unabhängigen Variablen gebildet, aber nur zur Bestimmung der Regressionskoeffizienten. Für die Prognose wird die unabhängige Variable nicht benützt.

Die Methode, wie sie hier beschrieben ist, hat in den Beispielen gute Resultate geliefert und zum Teil eine bessere Prognose ergeben als die klassische Methode. Aber auch wenn die Prognose nur wenig oder gar nicht verbessert wird, bleibt doch der wesentliche Vorteil, dass die t -Kenngrössen besser berechnet werden. Wie die Beispiele gezeigt haben, kann damit die Elimination der Regressoren allein anhand der Statistik erfolgen. Der Widerspruch mit den numerisch unwichtigen, aber statistisch signifikanten Regressoren verschwindet.

Wenn eine Autokorrelation der Residuen vorhanden ist, kann mit der klassischen Methode der kleinsten Quadrate immer noch ein gutes Modell berechnet werden. Bei der Auswahl der Regressoren ist der Ingenieur aber mehr auf seine Erfahrung angewiesen, da die berechneten t -Kenngrössen statistisch nicht mehr korrekt sind. Auch eine grosse Varianz der Regressionskoeffizienten muss in Kauf genommen werden, was eine weniger gute Prognose zur Folge haben kann. Der Fall mit täglichen Messwerten, wo die Autokorrelation sehr ausgeprägt ist, ist vielleicht nicht so häufig anzutreffen, wird

aber mit dem Trend zu automatischen Messungen in Zukunft eine vermehrte Rolle spielen.

5 Regressionsmodelle für Staumauern

5.1 Einflussgrössen

Die häufigste Anwendung der linearen Regression bei Staumauern ist die Überwachung von Verschiebungen. Die wichtigsten Einflussgrössen sind die Höhe des Wasserspiegels und die Temperatur. Wenn keine Temperaturmessungen zur Verfügung stehen, kann der Temperatureinfluss durch eine saisonale Funktion approximiert werden. Ebenfalls betrachtet werden Funktionen für irreversible Verschiebungen.

Neben diesen primären Einflüssen werden hier auch noch sekundäre Einflüsse betrachtet. Es handelt sich dabei um abgeleitete Grössen, Grössen die nicht direkt gemessen werden, sondern aus Messwerten berechnet werden. Betrachtet werden hier Temperaturen im Innern der Mauer, berechnet aus den Temperaturmessungen nahe der Oberfläche und Kriechen, berechnet aus den elastischen Verformungen aus Wasserdruck. Beide diese Grössen beinhalten einen Einfluss der gemessenen Grössen über eine gewisse Zeitdauer. Die Bestimmung von sekundären Einflussgrössen ist nicht der Hauptgegenstand dieser Untersuchung, jedoch ist ihre Auswahl ein wichtiges Thema, da sie oft zu Multikollinearität führen.

5.2 Polynome

Der Einfluss der Stauhöhe auf die Verschiebung kann im Allgemeinen mit einem Polynom von höchstens viertem Grade erfasst werden. Da die Stauhöhe gerne als Höhe über Meer angegeben wird, muss zuerst auf eine neue Referenz umgerechnet werden, da sonst die Potenzen zu grossen Zahlen führen würden. Oft wird die minimale Stauhöhe als Referenz verwendet.

Polynome tendieren dazu, linear abhängige Regressoren zu produzieren. Eine erste Massnahme zur Vermeidung des Problems ist die bereits angedeutete Verschiebung der Referenzlage. Besser als ein Minimalwert ist vom Standpunkt der Multikollinearität der Mittelwert als Referenz. Grössen mit dem Mittelwert als Referenz werden auch zentriert genannt. Die Zentrierung erfolgt bei den gemessenen Grössen vor dem Potenzieren. Die resultierenden Potenzen sind nicht mehr zentriert, wie anhand des quadratischen Terms unschwer festgestellt werden kann. In Bild 5-1 wird erläutert, wie die Zentrierung hilft, die lineare Abhängigkeit zwischen einer linearen und einer quadratischen Funktion zu vermeiden. Links sind die Daten nicht zentriert, sie können durch die lineare wie durch die quadratische Funktion approximiert werden. Rechts sind die Daten zentriert, die negativen Werte können klar den einzelnen Funktionen zugeteilt werden.

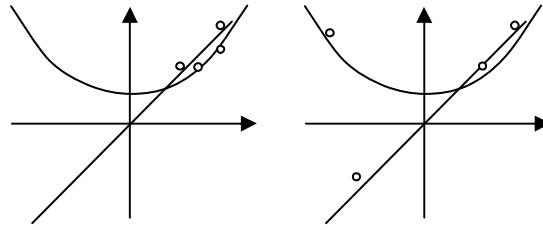


Bild 5-1: Lineare und quadratische Funktionen von zentrierten und nichtzentrierten Daten

Noch näher bei orthogonalen Regressoren sind Tschebyscheff-Polynome (Seber 1977, S. 215). Diese sind rekursiv definiert als

$$T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x). \quad (5.1)$$

Die ersten beiden Polynome sind

$$T_0(x) = 1 \quad \text{und} \quad T_1(x) = x. \quad (5.2)$$

Die weiteren Polynome ergeben sich aus der rekursiven Definition. Ein paar sind hier explizit angegeben:

$$\begin{aligned} T_2(x) &= 2x^2 - 1 \\ T_3(x) &= 4x^3 - 3x \\ T_4(x) &= 8x^4 - 8x^2 + 1 \\ T_5(x) &= 16x^5 - 20x^3 + 5x. \end{aligned} \quad (5.3)$$

Die Daten müssen so normiert werden, dass sie zwischen -1 und +1 liegen, also

$$x = \frac{2z - \max(z_i) - \min(z_i)}{\max(z_i) - \min(z_i)}. \quad (5.4)$$

Die Orthogonalität leitet sich aus der Tatsache ab, dass bei dieser Normierung die Transformation $x = \cos \theta$ möglich ist, und die Tschebyscheff-Polynome als $T_n(x) = \cos n\theta$ geschrieben werden können. Es ist dann

$$\int_{-1}^1 T_m(x) T_n(x) (1-x^2)^{-1/2} dx = \int_0^\pi \cos m\theta \cos n\theta d\theta = 0 \quad \text{für } m \neq n. \quad (5.5)$$

Wie gut dieses Resultat im diskreten Fall aussieht, hängt von der Verteilung der Daten ab. Am günstigsten ist es, wenn die Daten für eine gleichmäßige Verteilung von θ wie $x = \cos \theta$ verteilt sind, was für Stauhöhen fast idealerweise der Fall ist. Die Berechnung ist kaum aufwändiger als für die einfachen Polynome.

5.3 Periodische Einflüsse (Saisonale Einflüsse)

Falls Temperaturmessungen fehlen, kann der saisonale Einfluss durch eine Fourier-Reihe mit einer Jahresperiode approximiert werden. Normalerweise werden die Terme $\sin s$, $\cos s$, $\sin 2s$, $\cos 2s$ verwendet. Oft sieht man auch statt $\sin 2s$, $\cos 2s$ die äquivalenten Terme $\sin s \cos s = 1/2 \sin 2s$ und $\sin^2 s = 1/2(1 - \cos 2s)$. Dieser Ansatz minimiert den Aufwand bei der Auswertung von trigonometrischen Funktionen, ein Vorteil

der heute kaum mehr ins Gewicht fallen dürfte. Die Zeitvariable s ist so skaliert, dass sie in einem Jahr die Periode 2π durchläuft (365.25 Tage pro Jahr).

5.4 Wärmeleitung

Die Deformation einer Betonmauer hängt stark von der Temperaturverteilung ab. Meist wird die Temperatur aber nur an wenigen Punkten gemessen, oft in der Nähe der Oberfläche, und die Verteilung über die Dicke der Mauer ist nicht bekannt. Ohne eine aufwändige Wärmeleitungsberechnung zu machen, kann die Temperatur approximativ durch eine eindimensionale Berechnung bestimmt werden. Eine weitere Vereinfachung erhält man durch die Annahme, dass die Mauer so dick ist, dass sie als halbinendlich betrachtet werden kann. Diese Annahme ist erfüllt, wenn der Einfluss des einen Randes bis zum andern Rand soweit abnimmt, dass er vernachlässigbar ist.

Die eindimensionale Wärmeleitung kann durch die Gleichung

$$\frac{\partial T}{\partial t} = a \frac{\partial^2 T}{\partial z^2} \quad (5.6)$$

beschrieben werden. Dabei ist T die Temperatur, t die Zeit und z die Tiefe in die Mauer. Die Konstante a ist die Diffusionskonstante (thermal diffusivity). Sie kann ausgedrückt werden als

$$a = \frac{k}{c\rho} = \frac{\text{Wärmeleitfähigkeit}}{\text{spezifische Wärmekapazität} \times \text{Massendichte}}. \quad (5.7)$$

Ihr Wert für Beton ist ungefähr 0.05–0.15 m²/Tag (FERC 1999, S. 11-43).

Für den halbinendlichen Fall und eine harmonische Schwankung der Oberflächentemperatur $T = T_0 \sin \omega t$ ist die Lösung (Carslaw und Jaeger 1959, S.65)

$$T(z, t) = T_0 e^{-z/d} \sin(\omega t - z/d), \quad (5.8)$$

wobei d eine charakteristische Tiefe darstellt, gegeben durch

$$d = \sqrt{\frac{2a}{\omega}}. \quad (5.9)$$

Für jährliche Temperaturschwankungen und eine Diffusionskonstante von 0.1 m²/Tag ist $d = \sqrt{0.1 \cdot 2 \cdot 365 / 2\pi} = 3.4$ m. In dieser Tiefe klingen die Schwankungen auf $e^{-1} = 0.37$ ab und werden erst mit einer Verzögerung von $t = z/d / \omega = 365 / (2\pi) \approx 60$ Tagen festgestellt. Die Abklingfaktoren und Zeitverschiebungen bei verschiedenen Tiefen sind für jährliche Temperaturschwankungen in Tabelle 5-1 zusammengestellt. In einer Tiefe von 10 m spürt man nur noch 5% der jährlichen Aussentemperaturschwankungen, sodass die Annahme des halbinendlichen Gebietes für die meisten Mauern gerechtfertigt ist. Raschere Schwankungen nehmen mit der Tiefe noch schneller ab.

Tabelle 5-1: Abklingfaktoren und Zeitverschiebungen für $a = 0.1$ [m²/Tag]

z [m]	2	4	6	8	10
$e^{-z/d}$	0.56	0.31	0.17	0.10	0.05
$z/d/\omega$ [Tage]	34	68	102	136	170

Aufschlussreich ist auch eine Betrachtung der Temperaturverteilung zu verschiedenen Zeitpunkten. Bild 5-2 zeigt die Verteilung für jährliche Schwankungen. Es ist klar, dass die Temperaturen nicht einfach interpoliert werden können. Zur Zeit $t = 0$ z. B. ist die Temperatur in 2.5 m Tiefe nicht vernachlässigbar, obwohl sie am Rand und in 10 m Tiefe Null ist.

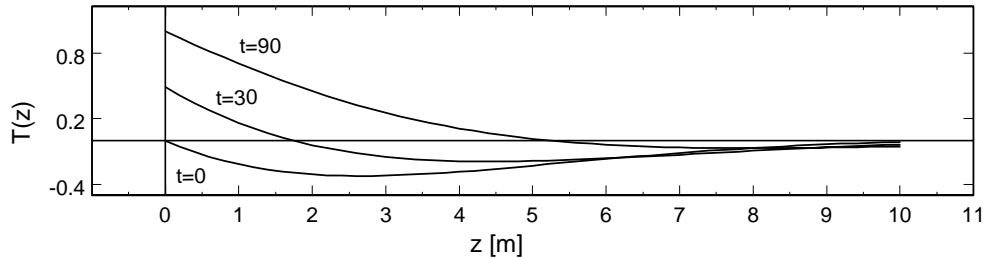


Bild 5-2: Temperaturverteilung im halbunendlichen Medium nach 0, 30 und 90 Tagen

Die Rede ist hier immer von Temperaturschwankungen, nicht von der Temperatur selber. Für die Berechnung wird ein konstanter Mittelwert von den gemessenen Temperaturen abgezogen, was keinen Einfluss auf die Regression hat.

Wenn man nicht nur die jährlichen Temperaturschwankungen berücksichtigen will, kann man die Temperatur in eine Fourier-Reihe zerlegen. Die andere Variante ist die Formulierung mit einem Konvolutionsintegral (Duhamel-Integral):

$$T(z, t) = \int_0^t T(z=0, \tau) g(t - \tau) d\tau. \quad (5.10)$$

Die Impulsantwort $g(t)$ für eine Tiefe z ist gegeben durch (Carslaw 1959, S.63)

$$g(t) = \frac{z}{2\sqrt{\pi a}} \frac{e^{-z^2/4at}}{t^{3/2}}. \quad (5.11)$$

Auch hier werden nur Schwankungen um die mittlere Temperatur berücksichtigt. Die Funktion $g(t)$ ist in Bild 5-3 für $z = 2, 4$ und 6 [m] Tiefe geplottet.

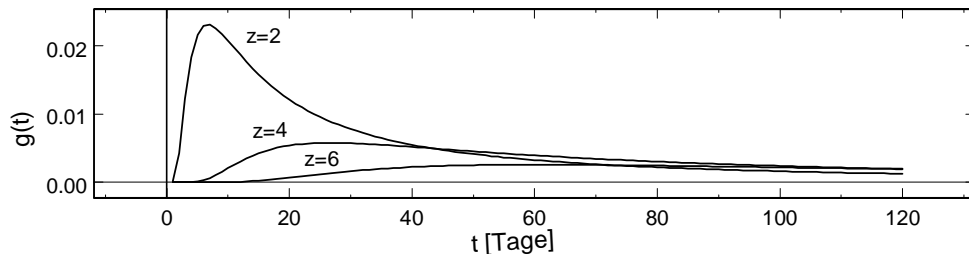


Bild 5-3: Funktion $g(t)$ für $a = 0.1$ [m²/Tag] und $z = 2, 4, 6$ [m]

Obwohl die Formulierung mit dem Konvolutionsintegral etwas komplizierter aussieht, ist sie doch die geeignete Form in diesem Zusammenhang. Vorausgesetzt die Temperaturen sind genügend fein diskretisiert, kann das Konvolutionsintegral mit der Trapezre-

gel einfach ausgewertet werden. Dabei können auch ungleichmässige Zeitschritte einfach berücksichtigt werden. (Eine Fourier-Transformation wäre in diesem Fall eher komplizierter). Von der Diskretisierung her sollte die Tiefe von $z = 2$ m nur für Zeitschritte von höchstens einer Woche und die Tiefe $z = 4$ m für Zeitschritte von höchstens einem Monat benützt werden.

Am Anfang der Berechnungsphase stehen keine Werte aus der Vergangenheit zur Verfügung. Der wirkliche Zustand stellt sich daher erst nach einer gewissen Anlaufzeit ein. Theoretisch kann die Regression erst nach dieser Anlaufzeit beginnen, aber damit vernachlässigt man viele Daten, die wichtiger sein können als eine genaue Temperatur, wie z.B. die Stauhöhe. Daher wird man als Kompromiss die Anlaufzeit eher kürzer wählen und in Kauf nehmen, dass am Anfang der Regression die berechneten Temperaturen leicht verfälscht sind. Eine andere Überlegung ist, wie weit zurück man den Einfluss der Vergangenheit berücksichtigen soll. Um die Berechnung zu beschleunigen, kann der Einfluss von Temperaturen, die weiter als z.B. ein Jahr zurückliegen, vernachlässigt werden.

Eine einfachere Alternative ist die Mittelbildung über eine gewisse Zeit in der Vergangenheit. Eine gewisse Schwierigkeit besteht bei ungleichmässigen Zeitschritten, da nicht immer die gleiche Anzahl Werte berücksichtigt werden muss und die Mittelungszeit nicht klar definiert ist. Die Berechnung mit dem Konvolutionsintegral kann in diesem Sinne als gewichtetes Mittel angesehen werden, bei dem die weiter zurück liegenden Temperaturen ein kleineres Gewicht haben. Da spielt es eine kleinere Rolle, wie weit zurück die Auswertung gemacht wird. Der Sachverhalt ist in Bild 5-4 schematisch dargestellt.

Es geht es hier nicht um eine genaue Berechnung, sondern um das Erfassen des wesentlichen Einflusses. In welcher Tiefe die Temperatur sich am meisten auf die Verformung auswirkt, hängt von der Geometrie und vom Beobachtungspunkt der Verschiebung ab. Grundsätzlich kann man eine Reihe von Temperaturverläufen in verschiedenen Punkten des Querschnitts berechnen und diese in das Modell aufnehmen. Bei zu vielen ähnlichen Regressoren ergeben sich aber wegen der Multikollinearität Schwierigkeiten bei der Variablenauswahl. Man sollte sich daher überlegen, welche Temperaturen physikalisch tatsächlich in Frage kommen und die Anzahl beschränken.

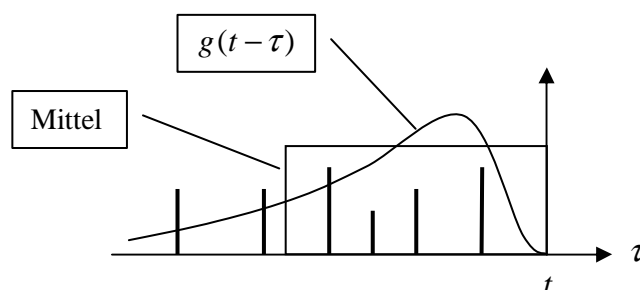


Bild 5-4: Vergleich Mittelbildung und Konvolutionsintegral

5.5 Kriechen

Kriechen kann wie die Wärmeleitung durch ein Konvolutionsintegral erfasst werden. Für die Kriechfunktion

$$\varphi(t) = \varphi_0(1 - e^{-\alpha t}) \quad (5.12)$$

ist die Verformung aus Kriechen

$$d_{cr}(t) = \alpha \varphi_0 \int_0^t d_{el}(\tau) e^{-\alpha(t-\tau)} d\tau. \quad (5.13)$$

Die Kriechkonstante α hat für Beton den Wert $\alpha = 0.01$ [1/Tag]. Die Konstante $\alpha \varphi_0$ spielt für die Regression keine Rolle und kann weggelassen werden. Die elastische Verformung d_{el} wird in der Regression durch ein Polynom repräsentiert. Daher muss für jeden Polynom-Regressor ein entsprechender Kriech-Regressor gebildet werden.

5.6 Irreversible Verformungen (Drift)

Bei der Modellierung von Langzeiteffekten gehen die Meinungen auseinander. Die einen erlauben nur zeitlich abnehmende Funktionen, damit eine zunehmende Verformung einer Mauer rechtzeitig aus den Abweichungen festgestellt werden kann, die anderen wählen auch zeitlich zunehmende Funktionen und untersuchen damit das Langzeitverhalten einer Mauer. Diese Studie folgt der ersten Philosophie und betrachtet nur zeitlich abnehmende Funktionen. Diese beeinflussen die Prognose nicht stark, helfen aber bei der Regression die Langzeiteffekte besser von den Kurzzeiteffekten zu trennen. Werden keine Regressoren für Langzeiteffekte berücksichtigt, besteht die Gefahr, dass die übrigen Regressoren verfälscht werden.

Für irreversible Verformungen werden meistens Exponentialfunktionen verwendet. Die Zeitskala in den Exponentialfunktionen $e^{-\eta t}$ spielt eine Rolle, da der Einfluss je nachdem über eine grössere oder kleinere Zeitspanne vorhanden ist. Oft wird in der Praxis die Zeit auf ein Jahr skaliert (365.25 Tage pro Jahr), aber es kommen grundsätzlich aber auch längere Perioden in Frage. Das Problem bei mehreren Exponentialfunktionen mit verschiedenen Zeitskalen ist, dass diese Regressoren zu einer starken Multikollinearität führen können. Als Beispiel sind in Bild 5-5 die Funktionen e^{-t} , $e^{-t/2}$ und $e^{-t/4}$ geplottet. Ebenfalls geplottet ist der Mittelwert von e^{-t} und $e^{-t/4}$, der der Funktion $e^{-t/2}$ zumindest in einem gewissen Bereich sehr ähnlich ist.

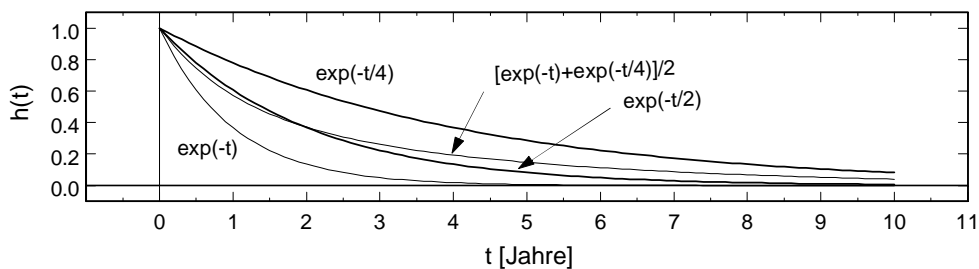


Bild 5-5: Vergleich von Exponentialfunktionen mit verschiedener Zeitskala

Wie sich die Multikollinearität negativ auf die Prognose auswirken kann, wird im folgenden einfachen Beispiel gezeigt. In Bild 5-6 sind die beiden Funktionen

$$f_1(t) = 0.5(1 + e^{-t} - 2e^{-t/4}) \quad \text{und} \quad f_2(t) = 1.06e^{-t} - 1.96e^{-t/2} + 0.9e^{-t/4} \quad (5.14)$$

aufgezeichnet. Beides sind Kombinationen von Exponentialfunktionen, die in den ersten 3 Jahren fast identisch verlaufen. Wird dieses Intervall für die Regression benützt, können beide Kombinationen gewählt werden, ohne grossen Unterschied in den Residuen. Der Unterschied macht sich erst in der Prognose, also nach 3 Jahren, bemerkbar. Obwohl nur Abklingende Funktionen verwendet wurden, erscheint $f_1(t)$ in den ersten 10 Jahren trotzdem als zunehmende Funktion. Das Problem ist hier nicht nur die Multikollinearität sondern auch die Verwendung von Langzeitfunktionen, die weit über die Regressionsperiode hinaus einflussreich sind und nur langsam abnehmen.

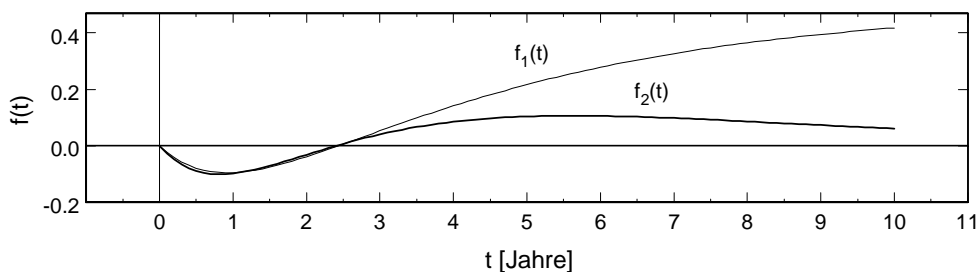


Bild 5-6: Günstige und ungünstige Kombinationen von Exponentialfunktionen

Exponentialfunktionen sind vor allem für die ersten Jahre nach dem Bau einer Mauer sinnvoll. Später erwartet man eigentlich keine irreversiblen Verformungen mehr, die diesen Verlauf haben. Sinnvoller scheinen da zeitlich lokale Funktionen, wie die Spline-Funktionen, die von Fanelli (2000) vorgeschlagen wurden. Diese verhindern, dass die Regressionskoeffizienten durch eine Vermischung von Kurzzeit- und Langzeiteinflüssen verfälscht werden. Auf die Prognose haben diese zeitlich begrenzten Funktionen keinen direkten Einfluss. Diese Art von Langzeitfunktionen wurde im Rahmen dieser Studie nicht weiter untersucht.

6 Anwendung auf Staumauer Schlegeis

In diesem und den beiden nächsten Kapiteln werden die bisher präsentierten Methoden anhand von drei Staumauern auf ihre praktische Anwendbarkeit hin überprüft. Dieses Kapitel behandelt die Staumauer Schlegeis, während die beiden nächsten Kapitel die Staumauer Sambuco und eine weitere, nicht näher spezifizierte Mauer untersuchen.

Die Beispiele sollen einerseits die Methoden illustrieren und andererseits die Schwierigkeiten aufzeigen, die entstehen können. Beim Vergleich der verschiedenen Modelle ist es wichtig, dass nicht nur die Regression sondern auch die Prognose beurteilt wird. Daher wurde immer nur ein Teil der Daten für die Regression verwendet, der andere Teil dient zur Überprüfung der Prognose.

6.1 Beschreibung

6.1.1 Überblick

Für die Schlegeis-Sperre wurden im Rahmen des 6. Benchmark Workshops 2001 in Salzburg (ICOLD 2002) Messreihen mit täglichen Werten über die Jahre 1992–2000 zur Verfügung gestellt. Ein Querschnitt der Mauer mit der Angabe der Messpunkte ist in Bild 6-1 gezeichnet.

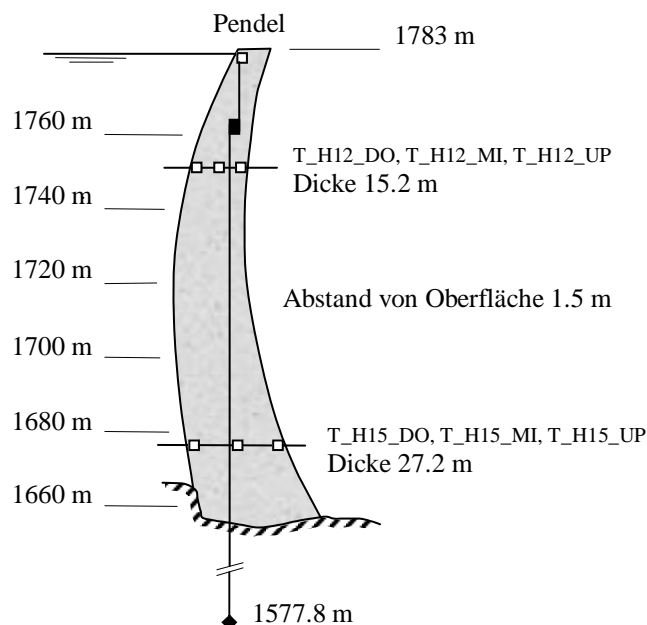


Bild 6-1: Querschnitt Schlegeis-Sperre mit Messpunkten

Zielgrösse ist die radiale Kronenverschiebung in Sperrenmitte. Die wichtigste Einflussgrösse ist die Stauhöhe. Die Betontemperaturen sind in 6 Punkten gegeben, auf zwei verschiedenen Messhorizonten (MH12 und MH15) je auf der Wasserseite (UP), der Luftseite (DO) und im Mauermitte (MI). Die Messpunkte an der Wasser- und der Luftseite haben einen Abstand von 1.5 m von der Oberfläche. Zusätzlich ist auch die Lufttemperatur gegeben. Die zahlreichen Messresultate erlauben es, eine Reihe von verschiedenen Modellen auszutesten.

Die Staumauer wurde zwischen 1969 und 1971 erbaut und der erste Vollstau war 1973. Für die betrachtete Periode sind daher keine irreversiblen Verschiebungen zu erwarten.

6.1.2 Messdaten

Die folgenden Plots geben eine Übersicht über die vorhandenen Daten. Die Zielgrösse ist die radiale Kronenverschiebung gemessen mit einem System von zwei Pendeln. Die Verschiebung ist für die letzten 2 Jahre (1999 und 2000) nicht gegeben, da die Aufgabe des Workshops darin bestand, diese Werte vorherzusagen. Der Verlauf der Kronenverschiebung ist in Bild 6-2 festgehalten.

Die wichtigste Einflussgrösse ist die Stauhöhe, gegeben in Bild 6-3. Weitere Grössen sind die Betontemperaturen auf den Messhorizonten MH12 und MH15 (siehe Skizze). Es sind je die Temperaturen auf der Wasserseite, der Luftseite und in Mauermitte gegeben. Die Temperaturen auf Messhorizont MH12 sind in Bild 6-4 und jene auf Messhorizont MH15 in Bild 6-5 gegeben. Am stärksten variieren die Temperaturen auf der Luftseite (untere Kurven). Auf der Wasserseite (obere Kurven) ist die Variation weniger gross, da die Temperaturschwankungen durch den Einfluss des Wassers verringert werden. Am wenigsten variieren die Betontemperaturen in der Mitte (mittlere Kurven) der Mauer. Bei der Temperatur T_{H12_MI} ist im Sommer 1994 eine kurzzeitige Spitze zu sehen, die wahrscheinlich auf ein Problem in der Messeinrichtung oder in der Datenübermittlung zurückzuführen ist. Der Einfluss dieser Betontemperatur ist aber eher gering und der Fehler nur kurzzeitig, sodass eine Handkorrektur nicht nötig erscheint. Es bleibt aber in den Residuen eine deutlich sichtbare Spitze bestehen. Als letztes ist in Bild 6-6 die Lufttemperatur geplottet.

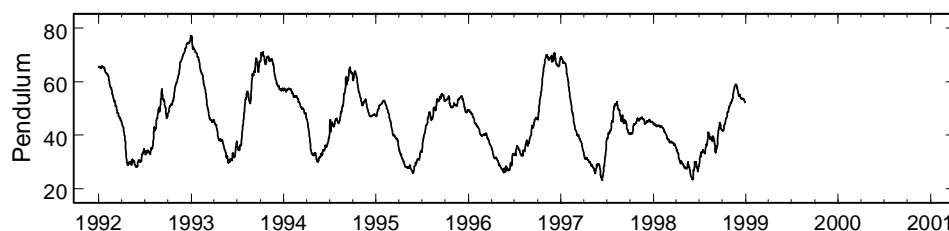


Bild 6-2: Kronenverschiebung

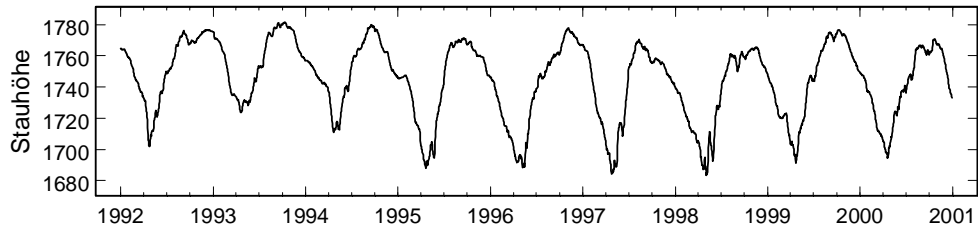


Bild 6-3: Stauhöhe

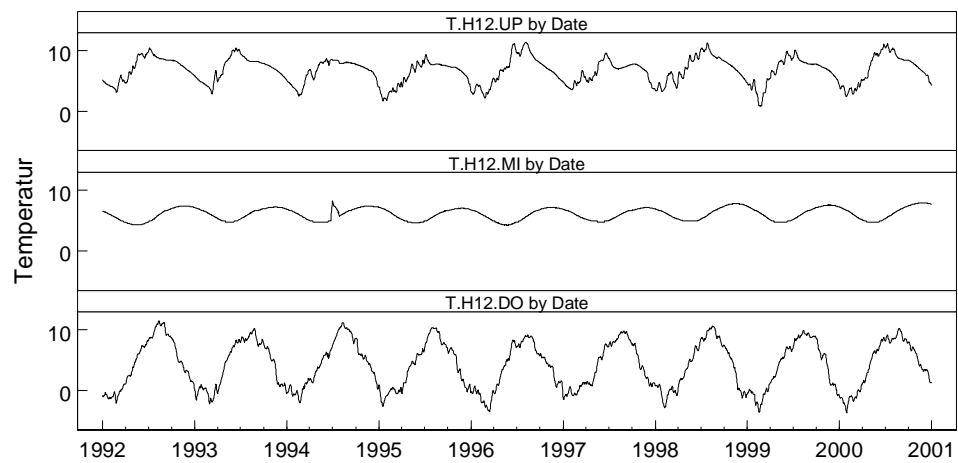


Bild 6-4: Betontemperaturen auf Messhorizont MH12: T_H12_UP, T_H12_MI, T_H12_DO

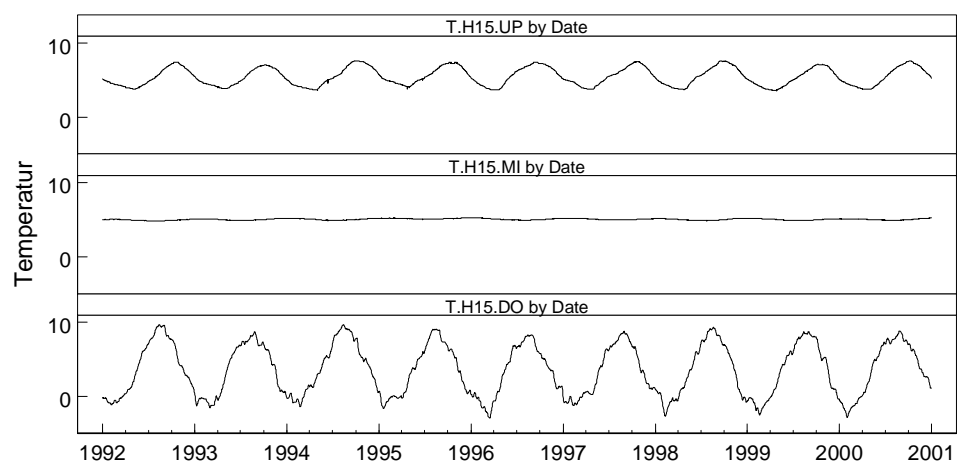


Bild 6-5: Betontemperaturen auf Messhorizont MH15: T_H12_UP, T_H12_MI, T_H12_DO

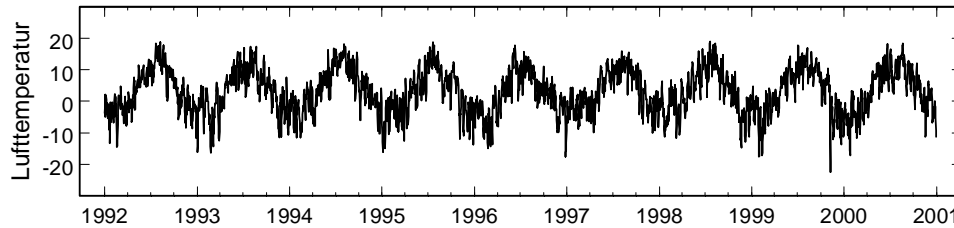


Bild 6-6: Lufttemperatur T_AIR_M

6.2 Anmerkung zu Tabellen und Plots

Die Tabellen mit den Resultaten wurden direkt aus dem Programm kopiert, mit dem die Berechnungen durchgeführt wurden. Auf eine Übersetzung der englischen Bezeichnungen wurde verzichtet. Der p -Wert für den globalen F -Test ist mit “signif F” bezeichnet, der Durbin-Watson-Kennwert mit “DW”, und “Observ” steht für die Anzahl Beobachtungen. “MS Res” ist die Quadratsumme der Residuen dividiert durch $n - k - 1$ (Anzahl Beobachtungen – Anzahl Regressoren inklusive Konstante) und ist ein unverzerrter Schätzwert für die Varianz der Fehler. Für den Vergleich der Regression mit der Prognose wurde aber die Quadratsumme direkt durch die Anzahl der Beobachtungen geteilt, da die erste Definition für die Prognose nicht sinnvoll ist. Die Unterschiede sind bei grösseren Messreihen minimal. Der Strich in “MS’ Res” soll den Unterschied markieren.

Bei den Residuenplots wurde auch ein Prognoseintervall eingezeichnet. Das Prognoseintervall berücksichtigt sowohl die Streuung der y -Variablen wie auch die Streuung der Regressionskoeffizienten. Das Intervall ist nicht konstant sondern abhängig vom Wert der Regressoren beim jeweiligen Prognosepunkt. In den meisten Fällen weicht dieses Prognoseintervall nicht wesentlich von einem Intervall ab, das nur eine Normalverteilung der y -Variablen berücksichtigt. Die 99.7 % entsprechen bei der Normalverteilung einem Intervall von ± 3 Standardabweichungen. Ebenfalls wurde in den Residuenplots markiert, wo die Prognose auf einer Extrapolation beruht. Es wird zwar nur angegeben, wo die Grenzen überschritten sind, aber diese Angabe kann trotzdem bei der Interpretation behilflich sein. Prognoseintervall und Extrapolation sind in Kapitel 3 beschrieben.

6.3 Klassisches Modell ohne Temperaturbeobachtungen

Das einfachste Modell verwendet nur Messungen des Wasserspiegels. Als Funktion wird ein Polynom vierten Grades gewählt. Um Probleme mit der Multikollinearität zu vermeiden, werden Tschebyscheff-Polynome verwendet. Gewöhnliche Polynome und der Einfluss der Normierung in werden später separat untersucht.

Der Temperatureinfluss wird über eine periodische Funktion über die Jahreszeiten berücksichtigt. Der Ansatz besteht aus Sinus- und Kosinustermen mit Perioden von einem ganzen und einem halben Jahr. Für eine zusätzliche inelastische Verformung wird eine exponentiell abfallende Funktion hinzugefügt, obwohl keine Langzeiteinflüsse zu erwarten sind. Gemäss gängiger Praxis werden zwei Zeitskalen benützt. Für die periodi-

schen (saisonalen) Funktionen ist $s = 2\pi d / 365$ und für die Langzeiteffekte ist $t = d / 365$, wobei d die Zeit in Tagen bezeichnet. (In der Berechnung wurden 365.25 Tage pro Jahr angenommen). Die Regression wird für die 4 Jahre 1992-1995 durchgeführt, die weitem 3 Jahre, 1996-1998, werden zur Überprüfung der Prognose benützt.

6.3.1 Methode der kleinsten Quadrate

Die Resultate der Regression mit der Methode der kleinsten Quadrate sind in Tabelle 6-1 zusammengestellt. Die p -Werte sind alle praktisch Null, ausser demjenigen für $\sin(2s)$, sodass statistisch gesehen alle andern Regressoren signifikant sind. Man sieht aber auch, dass einige der t -Kenngrössen deutlich kleiner sind als andere. Da die t -Kenngrössen ein direktes Mass sind für die Zunahme der Quadratsumme der Residuen, kann man also das Modell ohne allzu grosse Einbusse in der Genauigkeit verkleinern. Die schrittweise Elimination von 4 Regressoren führt zum Modell in Tabelle 6-2. Die Quadratsumme der Residuen (SS_{Res}) ist wie erwartet leicht angestiegen, aber nicht übermässig.

Tabelle 6-1: Ausgangsmodell mit saisonalen Funktionen

Regressor	Coefficient	Std Err	t-stat	p-value	VIF	Std Coeff
Constant	42.9811	0.09497	452.6	0	0	0
T1(W_LEVEL)	26.3648	0.2258	116.8	0	7.807	0.980803
T2(W_LEVEL)	7.58478	0.117	64.85	0	3.295	0.353895
T3(W_LEVEL)	1.53756	0.09902	15.53	2.041E-50	2.668	0.0762349
T4(W_LEVEL)	0.194493	0.07769	2.504	0.01241	2.15	0.0110352
$\sin(s)$	9.71599	0.1602	60.66	0	9.183	0.552544
$\cos(s)$	7.67137	0.06999	109.6	0	1.753	0.436267
$\sin(2s)$	-0.147612	0.07477	-1.974	0.04856	2.001	-0.00839465
$\cos(2s)$	-0.43713	0.06538	-6.686	3.267E-11	1.53	-0.0248594
$\exp(-t)$	-2.20444	0.1642	-13.42	8.282E-39	1.246	-0.0450443
SS Res	MS Res	R2	F	signif F	DW	Observ
2961.6	2.04108	0.98689	1.213E4	0	0.05	1461
Regression: 01/01/1992-31/12/1995			Prediction: 01/01/1996-31/12/1998			
SS Res	MS' Res		SS Res	MS' Res		
2961.6	2.02711		2530.82	2.30915		

Tabelle 6-2: Modell mit saisonalen Funktionen

Regressor	Coefficient	Std Err	t-stat	p-value	VIF	Std Coeff
Constant	42.8216	0.07663	558.8	0	0	0
T1(W_LEVEL)	25.3876	0.1751	145	0	4.035	0.94445
T2(W_LEVEL)	7.41841	0.1071	69.29	0	2.374	0.346132
T3(W_LEVEL)	1.91548	0.0836	22.91	1.826E-99	1.634	0.0949728
$\sin(s)$	8.95272	0.1243	72.04	0	4.752	0.509137
$\cos(s)$	7.88328	0.06571	120	0	1.328	0.448319
SS Res	MS Res	R2	F	signif F	DW	Observ
3454.34	2.37412	0.98471	1.874E4	0	0.04	1461
Regression: 01/01/1992-31/12/1995			Prediction: 01/01/1996-31/12/1998			
SS Res	MS' Res		SS Res	MS' Res		
3454.34	2.36437		3144.26	2.86885		

Die folgenden Bilder (Bild 6-7 und Bild 6-8) zeigen die Übereinstimmung der Regression mit der gemessenen Kronenverschiebung und die Residuen. Im Frühjahr 1996, 1997 und 1998 wird eine Extrapolation angezeigt, d.h. die Regressoren bewegen sich dort ausserhalb des Bereiches, der für die Regression verwendet wurde. Eine Betrachtung von Bild 6-3 zeigt, dass die gemessene Stauhöhe dort besonders niedrig war. Bild 6-9 zeigt, dass die Annahme einer Normalverteilung erfüllt ist, und dass die Residuen unabhängig von der Regressionsvariablen (berechnete Verschiebung) sind. Insbesondere ist die Streuung mehr oder weniger konstant. Die Residuen sind aber zeitlich voneinander abhängig, wie das Korrelogramm (Bild 6-10) zeigt.

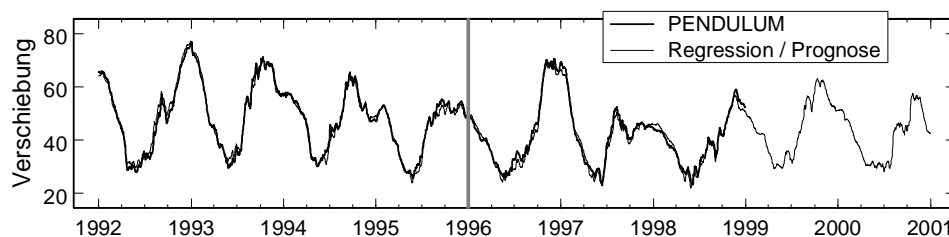


Bild 6-7: Regression, Modell mit saisonalen Funktionen

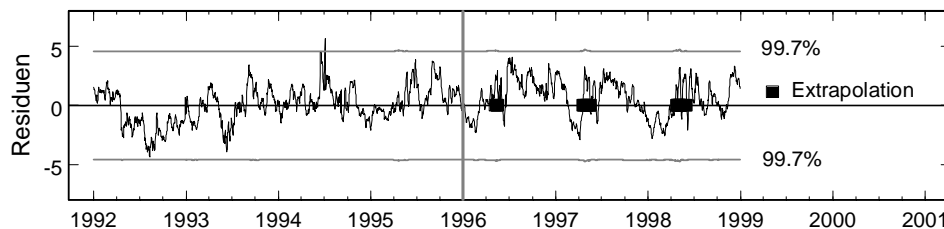


Bild 6-8: Residuen, Modell mit saisonalen Funktionen

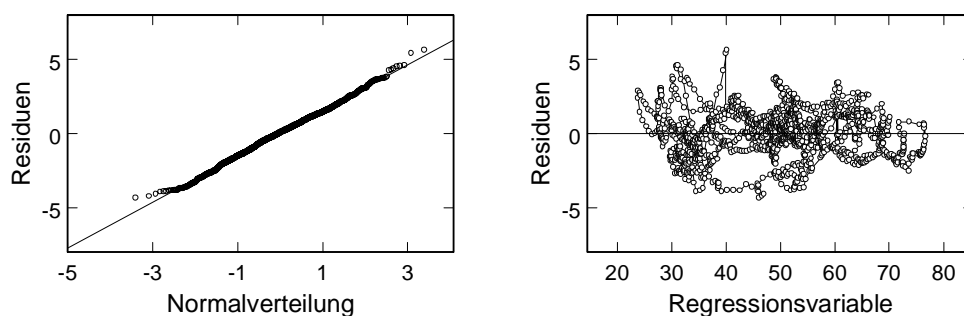


Bild 6-9: Verteilung der Residuen, Modell mit saisonalen Funktionen

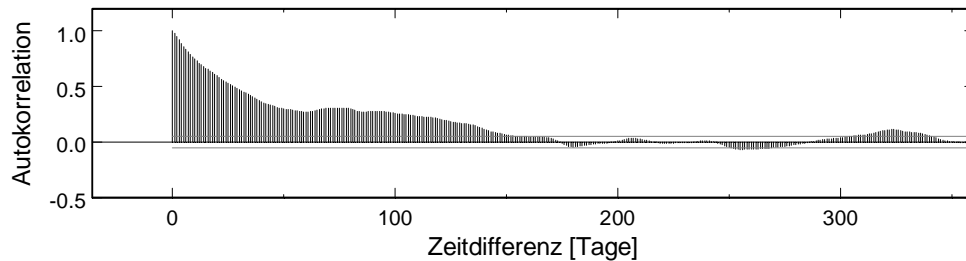


Bild 6-10: Korrelogramm, Modell mit saisonalen Funktionen

6.3.2 Autokorrelation der Fehler

Wie aus dem Korrelogramm ersichtlich ist, und wie der Durbin-Watson-Kennwert (DW) anzeigt, sind die Fehler autokorreliert. Der berechnete Wert von 0.04 liegt weit unter dem unteren 5%-Grenzwert von $d_L = 1.93$. Die Tabellen gehen meist nur bis 100 Beobachtungen, aber der Wert für 1461 Beobachtungen kann nach der Näherung in Durbin und Watson (1950, S.427) berechnet werden. Die Autokorrelation dürfte für diese Art von Anwendung typisch sein, insbesondere bei einem Zeitschritt von nur einem Tag. In dieser kurzen Zeit zwischen zwei Messungen können sich die Beobachtungen nur wenig ändern, und so ändern sich auch die Residuen nicht zufällig, sondern sind korreliert. Dies ist auch aus Bild 6-9 rechts ersichtlich, wo die Residuen durch eine Linie verbunden sind.

Mit der Autokorrelation der Fehler ist eine wichtige Annahme der Methode der kleinsten Quadrate verletzt. Die Theorie besagt, dass die Regressionskoeffizienten trotzdem richtig geschätzt werden (unverzerrt), dass aber ihre Varianz und damit die t -Kenngrößen falsch (typischerweise zu hoch) berechnet werden. Dies erklärt, warum im Beispiel auch weniger wichtige Regressoren statistisch als signifikant ausgegeben wurden. Ein weiterer Nachteil ist, dass die berechneten Regressionskoeffizienten nicht mehr effektiv sind, d.h. nicht mehr minimale Varianz haben. Die minimale Varianz der Regressionskoeffizienten ist aber eine wichtige Eigenschaft für die Prognose.

Die Autokorrelation der Fehler kann mit der verallgemeinerten Methode der kleinsten Quadrate berücksichtigt werden. Für eine Autokorrelation erster Ordnung kann man die Regressoren durch eine verallgemeinerte Differenzenbildung transformieren, und die gewöhnliche Methode der kleinsten Quadrate anwenden, wie dies in Kapitel 4 beschrieben wurde. Für das betrachtete Beispiel liefert die Formel nach Pesaran und Slater (1980, S. 37) den Wert $\rho = 1$. Eine Berechnung mit allen Regressoren führt zu den Werten in Tabelle 6-3. In dieser Tabelle beziehen sich alle Werte, mit wenigen Ausnahmen, auf die transformierten Variablen (abhängige und unabhängige), nicht auf die ursprünglichen Variablen. Die Regressionskoeffizienten und die t -Kenngrößen bleiben allerdings bei der Rücktransformation unverändert und gelten somit auch für das ursprüngliche Modell. Da $\rho = 1$ ist, fällt die Konstante für die transformierten Variablen weg und muss für die ursprünglichen Variablen so bestimmt werden, dass die Regressionsfunktion durch den Schwerpunkt geht. Das Bestimmtheitsmass und die Quadratsumme der Residuen beziehen sich ebenfalls auf die transformierten Variablen und können nicht direkt mit den Werten aus der klassischen Methode verglichen werden. In

den untersten Zeilen sind Quadratsummen für das ursprüngliche Modell angegeben, damit diese direkt mit andern Modellen verglichen werden können.

Die drei letzten Regressoren sind statistisch nicht signifikant (p -Werte über 1 %) und können deshalb weggelassen werden. Man kann sich also hier allein auf die Statistik abstellen. Im Gegensatz zur gewöhnlichen Methode der kleinsten Quadrate, bleibt hier das Polynom 4-ten Grades im Modell. Die t -Kenngrößen sind kleiner, was nicht zeigt, dass dieses Modell schlechter ist, sondern dass die früheren Werte wegen falschen Annahmen (unkorrelierte Fehler) falsch berechnet wurden. Der Durbin-Watson-Kennwert ist viel grösser geworden, wenn auch der Grenzwert $d_L = 1.93$ noch erreicht ist. Das reduzierte Modell ist in Tabelle 6-4 angegeben.

Interessant ist auch, dass das Modell ohne nicht-signifikante Regressoren sich besser verhält, als jenes mit allen Regressoren. Das Problem ist, dass die Exponentialfunktion in diesem Beispiel mit der verallgemeinerten Methode stark überbetont wird. Sobald sie eliminiert wird, verbessert sich das Verhalten.

Tabelle 6-3 : Ausgangsmodell mit saisonalen Funktionen, $\rho=1$

Regressor	Coefficient	StdErr	t-stat	p-value	VIF	Std Coeff
T1(W_LEVEL)	23.5991	0.6275	37.61	1.034E-216	1.54	0.701308
T2(W_LEVEL)	7.47164	0.2643	28.27	1.914E-140	1.131	0.451687
T3(W_LEVEL)	1.78689	0.1792	9.974	1.043E-22	1.106	0.157595
T4(W_LEVEL)	0.828646	0.1293	6.411	1.95E-10	1.051	0.0987273
sin(s)	7.34869	0.8059	9.119	2.459E-19	1.447	0.16477
cos(s)	8.16899	0.6888	11.86	4.972E-31	1.058	0.183288
sin(2s)	0.312142	0.3606	0.8657	0.3868	1.158	0.0139969
cos(2s)	-0.196634	0.3512	-0.5599	0.5756	1.1	-0.00882346
exp(-t)	5.75121	8.571	0.671	0.5023	0.5344	0.00736948
SS Res	MS Res	R2	F	signif F	DW	Observ
140.724	0.0969171	0.67228	330.9	0	1.70	1462
Regression: 01/01/1992-31/12/1995			Prediction: 01/01/1996-31/12/1998			
SS Res	MS' Res		SS Res	MS' Res		
7893.65	5.40291		6033.44	5.50496		

Tabelle 6-4: Modell mit saisonalen Funktionen, $\rho=1$

Regressor	Coefficient	Std Err	t-stat	p-value	VIF	Std Coeff
T1(W_LEVEL)	23.756	0.587	40.47	1.592E-240	1.35	0.705971
T2(W_LEVEL)	7.50889	0.2607	28.8	9.518E-145	1.102	0.45394
T3(W_LEVEL)	1.77894	0.1715	10.37	2.316E-24	1.015	0.156895
T4(W_LEVEL)	0.821124	0.1288	6.375	2.458E-10	1.045	0.0978311
sin(s)	7.45615	0.7926	9.407	1.91E-20	1.401	0.167179
cos(s)	8.20777	0.6791	12.09	4.092E-32	1.03	0.184158
SS Res	MS Res	R2	F	signif F	DW	Observ
140.871	0.0968185	0.67193	496.8	0	1.70	1462
Regression: 01/01/1992-31/12/1995			Prediction: 01/01/1996-31/12/1998			
SS Res	MS' Res		SS Res	MS' Res		
3705	2.53594		3114.35	2.84156		

Der Vergleich zur gewöhnlichen Methode ist in Tabelle 6-5 zusammengestellt. Die Spalte "Reg" gibt die Anzahl der Regressoren an. Die Quadratsummen sind in beiden Fällen praktisch gleich. Einen besseren Vergleich bietet die Zusammenstellung der Residuenplots am Ende dieses Kapitels.

Tabelle 6-5: Vergleich gewöhnliche und verallgemeinerte Methode (Tabelle 6-2 und Tabelle 6-4)

		Regression: 01/01/1992-31/12/1995		Prediction: 01/01/1996-31/12/1998	
	Reg	SS Res	MS' Res	SS Res	MS' Res
$\rho=0$	5	3454.34	2.36437	3144.26	2.86885
$\rho=1$	6	3705	2.53594	3114.35	2.84156

Bild 6-11 zeigt die Residuen mit dem neuen Modell. Sie sind praktisch identisch zu denjenigen in Bild 6-8. Die Folgenden Bilder dienen zur Überprüfung der Annahmen über die Verteilung der Fehler. Sie basieren auf den Residuen des transformierten Problems, die in Bild 6-12 geplottet sind. Es sind diese Residuen, die die Annahmen der Methode der kleinsten Quadrate erfüllen müssen. Auffällig sind die paar Ausreisser, besonders jene Mitte 1994. Die Normalverteilung und die Verteilung in Abhängigkeit der berechneten Verschiebung sind in Bild 6-13 festgehalten. Abgesehen von den Ausreissern sind die Annahmen der Normalverteilung und der konstanten Varianz erfüllt. Bild 6-14 zeigt das Korrelogramm. Wie man sieht, ist die Autokorrelation im transformierten Modell verschwunden und die Annahmen für die gewöhnliche Methode der kleinsten Quadrate sind erfüllt.

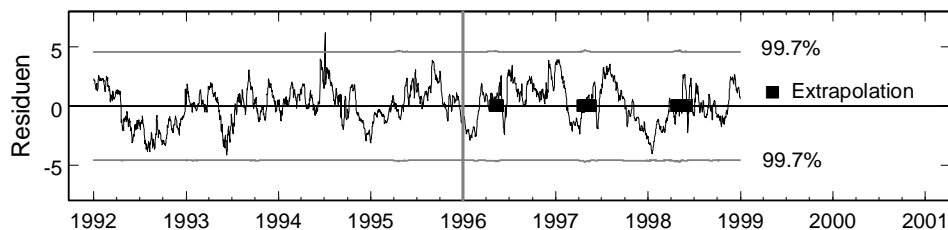


Bild 6-11 Residuen, Modell mit saisonalen Funktionen, $\rho=1$

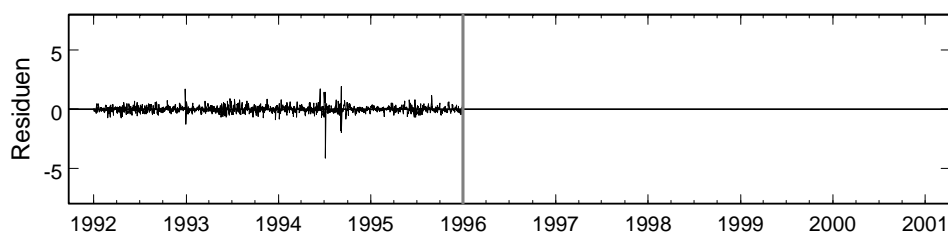


Bild 6-12: Residuen im transformierten Problem, Modell mit saisonalen Funktionen, $\rho=1$

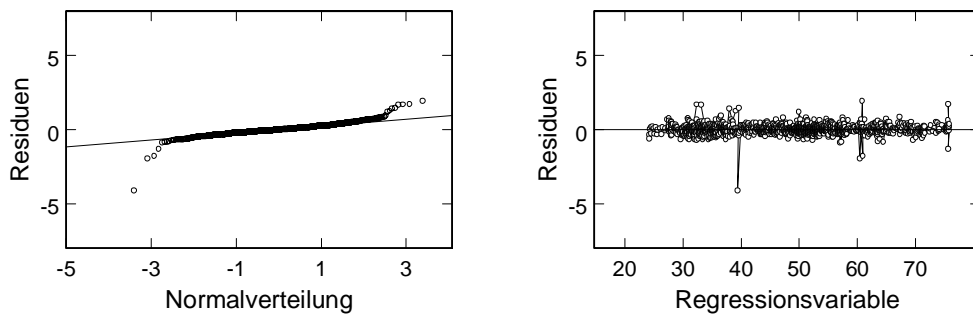


Bild 6-13: Verteilung der Residuen im transformierten Problem, Modell mit saisonalen Funktionen, $\rho=1$

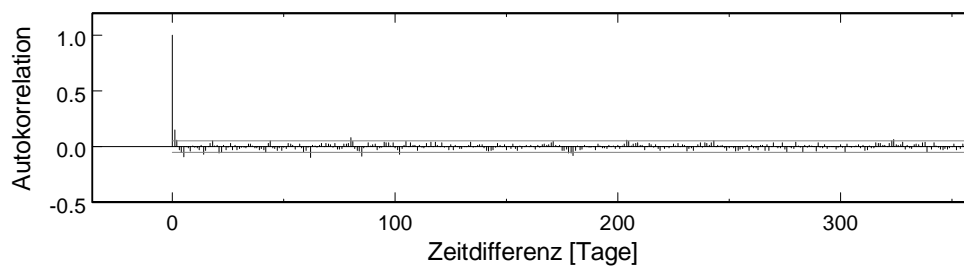


Bild 6-14: Korrelogramm im transformierten Problem, Modell mit saisonalen Funktionen, $\rho=1$

6.3.3 Abhängigkeit von der Regressionsperiode

Die Resultate einer Regression können stark von der Regressionsperiode abhängen. Die folgenden Resultate wurden mit einer Regressionsperiode von 3 statt 4 Jahren berechnet (1.1.1993 – 31.12.95). Tabelle 6-6 zeigt die Zusammenfassung der klassischen und der verallgemeinerten Methode. Auffallend ist, dass in beiden Fällen sowohl die Regression wie auch die Prognose viel besser ausfallen als bei der Regression über 4 Jahre (Tabelle 6-5). Dass die Regression bei einem kleineren Intervall zu kleineren Residuen führt, ist intuitiv verständlich, dass aber auch die Prognose besser wird, ist nicht unbedingt zu erwarten. Obwohl theoretisch die verallgemeinerte Methode eine effizientere Schätzung der Regressionsparameter liefern sollte, ist in diesem Beispiel die Prognose mit der klassischen Methode besser ausgefallen.

Das Beispiel zeigt auch klar, dass man bei dieser Art von Berechnungen eine grosse Streuung der Resultate erwarten muss. Dies sollte man sich vor Augen halten, wenn man im Folgenden kompliziertere Modelle mit diesem einfachen Modell vergleicht.

Tabelle 6-6: Modelle mit saisonalen Funktionen, Regressionsperiode 3 Jahre

		Regression: 01/01/1993-31/12/1995		Prediction: 31/12/1994-31/12/1998	
	Reg	SS Res	MS' Res	SS Res	MS' Res
$\rho = 0$	5	1734.96	1.58444	2345.45	2.14001
$\rho = 1$	6	2031.09	1.85487	2933.89	2.67691

6.3.4 Polynome und Skalierung

Eine Frage von mehr praktischer Bedeutung ist die der verwendeten Polynome für die Stauhöhe und deren Skalierung. Wie im theoretischen Teil beschrieben wurde, kann mit Tschebyscheff-Polynomen die beste Orthogonalität der Regressoren erreicht werden. Bei einfachen Polynomen ergibt sich die beste Unabhängigkeit mit einer Zentrierung, d.h. mit einem Nullpunkt, der dem Mittelwert entspricht. Etwas weniger gut ist ein Nullpunkt der Stauhöhe beim Minimalwert. Unbrauchbar ist die Stauhöhe direkt ab Meereshöhe gemessen, was aber allgemein bekannt sein dürfte und hier nur als Illustration verwendet wird. Eine Skalierung der Höhe hat aber auf die Orthogonalität keinen Einfluss und bringt keine numerischen Vorteile.

Die verschiedenen Polynom-Ansätze werden im Folgenden untersucht. Alle Vergleiche werden mit dem gleichen einfachen Modell für eine Regressionsperiode von 1992 bis 1995 durchgeführt. Tabelle 6-7 vergleicht verschiedene Annahmen für den Nullpunkt. Obwohl alle Fälle zu den gleichen Quadratsummen der Residuen führen, sind doch die VIF-Werte ganz unterschiedlich. Hohe VIF-Werte bedeuten gegenseitige Abhängigkeiten mit all ihren Nachteilen. Wie bereits gesagt, sind die VIF-Werte am grössten bei Polynomen mit absoluten Stauhöhen. Besser sind Polynome mit Stauhöhen von einem Minimalwert aus gemessen und am besten sind Polynome mit zentrierten Stauhöhen. Tabelle 6-8 vergleicht die entsprechenden Polynome mit einer zusätzlichen Skalierung, sodass grosse Zahlen vermieden werden. Obwohl es im Allgemeinen ein guter Grundsatz ist, Messwerte vor der Berechnung zu skalieren, sieht man aus den Vergleichen, dass die Skalierung hier keinen Einfluss auf die Orthogonalität hat. Tabelle 6-9 schliesslich zeigt die entsprechenden Werte für Tschebyscheff-Polynome. Es ist klar ersichtlich, dass hier die kleinsten VIF-Werte vorhanden sind, und somit die Orthogonalität am besten ist.

Tabelle 6-7: Verschiedene Nullpunkte für Stauhöhe, $H_{\min} = 1653$, $H_{\text{mean}} = 1751.3$

	H		H - H_{\min}		H - H_{mean}	
Regressor	Coefficient	VIF	Coefficient	VIF	Coefficient	VIF
Constant	5.60927E6	0	30.9382	0	44.6232	0
h	-13117	2.404E12	-0.582745	2.623E4	0.681131	17.73
h ²	11.5053	2.176E13	0.0155166	3.118E5	0.0100218	8.29
h ³	-0.00448627	2.189E13	-0.000147637	4.272E5	0.000110371	59.66
h ⁴	6.56176E-7	2.447E12	6.56176E-7	6.583E4	6.56176E-7	47.72
sin(s)	8.8246	4.975	8.8246	4.975	8.8246	4.975
cos(s)	7.89968	1.332	7.89968	1.332	7.89968	1.332
	SS Res	MS Res	SS Res	MS Res	SS Res	MS Res
	3400.49	2.33871	3400.49	2.33871	3400.49	2.33871

Tabelle 6-8: Verschiedene Skalierungen für Stauhöhe, $H_{\min} = 1653$, $H_{\max} = 1782$, $H_{\text{mean}} = 1751.3$

	H / H_{\max}		$(H - H_{\min}) / (H_{\max} - H_{\min})$		$(H - H_{\text{mean}}) / (H_{\max} - H_{\min})$	
Regressor	Coefficient	VIF	Coefficient	VIF	Coefficient	VIF
Constant	5.60927E6	0	30.9382	0	44.6232	0
h	-2.33744E7	2.404E12	-75.1735	2.623E4	87.8652	17.73
h^2	3.6535E7	2.176E13	258.208	3.118E5	166.77	8.29
h^3	-2.53865E7	2.189E13	-316.923	4.272E5	236.927	59.66
h^4	6.61672E6	2.447E12	181.704	6.583E4	181.704	47.72
sin(s)	8.8246	4.975	8.8246	4.975	8.8246	4.975
cos(s)	7.89968	1.332	7.89968	1.332	7.89968	1.332
	SS Res	MS Res	SS Res	MS Res	SS Res	MS Res
	3400.49	2.33871	3400.49	2.33871	3400.49	2.33871

Tabelle 6-9: Tschebyscheff Polynome

	Tschebyscheff	
Regressor	Coefficient	VIF
Constant	42.9801	0
h	25.1616	4.331
h^2	7.70971	3.149
h^3	1.71019	2.069
h^4	0.391123	2.066
sin(s)	8.8246	4.975
cos(s)	7.89968	1.332
	SS Res	MS Res
	3400.49	2.33871

Ein Nachteil von gegenseitig stark abhängigen Regressoren ist die schlechte numerische Konditionierung, die zu ungenauen Resultaten führen kann. Die vorangegangenen Beispiele zeigen aber keine solchen Probleme. Sie wurden mit doppelter Genauigkeit und mit der QR-Zerlegung berechnet (siehe Anhang). Diese Methode ist viel weniger empfindlich als die Normalgleichungen. Zum Vergleich wurden in Tabelle 6-10 einige Berechnungen mit Excel wiederholt. Es ist zwar nicht bekannt mit welcher Methode und welcher Genauigkeit Excel arbeitet, aber man sieht, dass bei sehr grossen VIF-Werten grosse Ungenauigkeiten entstehen. Bei Polynomen mit absoluten Stauhöhen sind zwar die Quadratsummen der Residuen nur unwesentlich grösser, aber die Regressionskoeffizienten sind total verschieden von denjenigen, die mit der QR-Zerlegung berechnet wurden. Daran ändert auch eine Skalierung nichts. Erst die Verschiebung des Nullpunkts löst das Problem. Bei der Verwendung der Minimalhöhe als Nullpunkt, die zwar bezüglich der Orthogonalität noch nicht optimal ist, sind die Resultate von Excel und der Berechnung mit der QR-Zerlegung praktisch identisch, sodass dieser Ansatz vom numerischen Standpunkt aus absolut vertretbar ist.

Tabelle 6-10: Vergleich QR-Zerlegung und Excel für verschiedene Normierungen

	H		H / H _{max}		H - H _{min}	
	QR	Excel			QR	Excel
Regressor	Coefficient	Coefficient	Coefficient	Coefficient	Coefficient	Coefficient
Constant	5.60927E6	-135215	5.60927E6	157636.8	30.9382	30.93824
h	-13117	108.6492	-2.33744E7	-1008494	-0.582745	-0.58274
h ²	11.5053	0.087778	3.6535E7	2128966	0.0155166	0.015517
h ³	-0.00448627	-0.00011	-2.53865E7	-1865492	-0.000147637	-0.00015
h ⁴	6.56176E-7	2.61E-08	6.61672E6	587461.3	6.56176E-7	6.56E-07
sin(s)	8.8246	8.946749	8.8246	8.939593	8.8246	8.824626
cos(s)	7.89968	7.884414	7.89968	7.88569	7.89968	7.899659
	SS Res	SS Res	SS Res	SS Res	SS Res	SS Res
	3400.49	3450.152	3400.49	3445.216	3400.49	3400.494

Grundsätzlich muss beachtet werden, dass die hier besprochenen Kollinearitäten nicht essentielle Kollinearitäten sind, sondern mit einer geeigneten Wahl der Polynome vermieden werden können. Die Folgen dieser Kollinearitäten sind nur numerischer Art, sonst sind die Modelle äquivalent. Da die beteiligten Regressoren alle aus den gleichen gemessenen Variablen hergeleitet sind, ist garantiert, dass ihre gegenseitige Abhängigkeit auch bei einer Prognose bestehen bleibt, und somit auch die Voraussagen bei allen Modellen gleich sind. Kleine VIF-Werte erweisen sich erst dann als Vorteil, wenn auch andere Regressoren voneinander abhängig sind. Im nächsten Beispiel sind zwei Betontemperaturen stark korreliert und ergeben entsprechend hohe VIF-Werte. Haben die Polynome kleine VIF-Werte, so sieht man rasch, dass die beiden Temperaturen korreliert sind. Zeigen aber auch die Polynome hohe VIF-Werte, so wird die Angelegenheit bald einmal unübersichtlich.

6.4 Modell mit Betontemperaturen

Weitere Regressoren, die man einbeziehen kann, sind die Betontemperaturen. Im Modell werden wieder 4 Tschebyscheff-Polynome und die Exponentialfunktion berücksichtigt. Die saisonalen Einflüsse werden durch die 6 Betontemperaturen abgedeckt, sodass man auf die Sinus- und Kosinus-Ansätze verzichten kann. Auch hier wird wieder eine Regressionsperiode von 4 Jahren gewählt und eine Prognoseperiode von 3 Jahren. Die Resultate sind in Tabelle 6-11 zusammengestellt. Zwei Regressoren, T_H12_MI und exp(-t), haben grosse p -Werte (über 1 %) und sind daher statistisch nicht signifikant. Zwei Regressoren, T_H12_DO und T_H15_DO, haben grosse VIF-Werte (über 10). Es sind dies die beiden Temperaturen auf der Luftseite, die einen sehr ähnlichen Verlauf zeigen (Bild 6-4 und Bild 6-5). Da die beiden Regressoren stark korreliert sind, ist ihre t -Kenngrösse nicht massgebend für ihre Wichtigkeit im Endmodell, sondern nur für den nächsten Eliminationsschritt. Zusätzlich ist der Durbin-Watson-Kennwert sehr klein (viel näher bei 0 als bei 2), sodass hier zusätzlich eine Autokorrelation der Fehler vorhanden ist und damit die t -Kenngrössen inkorrekt sind.

Tabelle 6-11: Ausgangsmodell mit Betontemperaturen

Regressor	Coefficient	Std Err	t-stat	p-value	VIF	Std Coeff
Constant	34.1141	3.685	9.256	7.365E-20	0	0
T1(W_LEVEL)	25.946	0.1735	149.6	0	5.317	0.965226
T2(W_LEVEL)	7.14064	0.1086	65.77	0	3.277	0.333172
T3(W_LEVEL)	2.25811	0.08638	26.14	1.022E-123	2.342	0.111961
T4(W_LEVEL)	0.301038	0.07202	4.18	3.093E-5	2.132	0.0170804
T_H12_UP	-0.430218	0.0422	-10.19	1.293E-23	5.41	-0.0663562
T_H12_MI	-0.0274321	0.1121	-0.2448	0.8067	10.23	-0.00219128
T_H12_DO	-0.941022	0.09217	-10.21	1.107E-23	104.8	-0.292502
T_H15_UP	-1.0772	0.09906	-10.87	1.567E-26	11.61	-0.103698
T_H15_MI	4.931	0.687	7.178	1.129E-12	6.455	0.0510364
T_H15_DO	-0.864018	0.1076	-8.028	2.033E-15	114.3	-0.24015
exp(-t)	0.0413552	0.2557	0.1618	0.8715	3.485	0.000845028
SS Res	MS Res	R2	F	signif F	DW	Observ
2563.45	1.76912	0.98865	1.148E4	0	0.07	1461
Regression: 01/01/1992-31/12/1995			Prediction: 01/01/1996-31/12/1998			
SS Res	MS' Res	SS Res		MS' Res		
2563.45	1.75459	3656.08		3.33584		

Es gibt nun zwei Wege, das Problem weiter zu verfolgen. Der eine ist der, dass man die Autokorrelation der Fehler zunächst nicht beachtet und die gewöhnliche Methode der kleinsten Quadrate benützt, um die Regressionskoeffizienten zu berechnen. Man weiss ja, dass diese korrekt berechnet werden. Für die Elimination der Regressoren stützt man sich auf die t -Kenngrössen, die zwar statistisch falsch sind, deren Werte aber ein direktes Mass dafür sind, um wie viel sich bei der Elimination die Quadratsumme der Residuen vergrössern wird. Das Problem ist dann aber die Multikollinearität, angezeigt durch die hohen VIF-Werte. Es ist instruktiv dieses Problem mit der Principal-Component-Regression genauer zu untersuchen.

Der zweite Weg ist der, dass man zuerst die Autokorrelation der Fehler beseitigt, indem man die verallgemeinerte Methode der kleinsten Quadrate verwendet. Dies geht in diesem Fall nicht ganz ohne Schwierigkeiten. Wie man sehen wird, konvergiert das Verfahren zunächst nicht zu einer brauchbaren Lösung. Es bleiben sinusförmige Residuen mit einer grossen Amplitude. Erst eine Optimierung des Korrelationsparameters ρ von Hand, führt schliesslich zum Ziel. Mit diesem Verfahren verschwinden die hohen VIF-Werte und man kann direkt den t -Test benützen um die unbedeutenden Regressoren zu eliminieren.

Es werden hier beide Wege gezeigt, obwohl der erste vom theoretischen Standpunkt aus nicht ganz korrekt ist. Trotzdem wird er aus praktischen Überlegungen vielleicht vorgezogen und er führt zu einer annehmbaren Lösung. Man sollte sich nur im Klaren sein, dass die berechneten t -Kenngrössen statistisch nicht korrekt sind.

6.4.1 Beseitigung der Multikollinearität durch Regressor-Elimination

Es wird also zunächst das obige Modell weiter untersucht, ohne auf das Problem der Autokorrelation der Fehler einzugehen. Wie schon früher beobachtet, verlaufen die beiden Temperaturen mit hohen VIF-Werten (T_H12_DO und T_H15_DO) praktisch parallel. Dies bestätigt auch ihr Korrelationskoeffizient von 0.99. Da beide ungefähr die

gleiche Information enthalten, ist die eine Strategie, nur einen von beiden Regressoren im Modell zu behalten. Die Elimination von T_H15_DO führt zum Modell in Tabelle 6-12. Der verbleibende Regressor T_H12_DO erhält einen Koeffizienten von -1.63306 , was ungefähr der Summe der beiden früheren Koeffizienten (-0.941022 und -0.864018) entspricht. Dies bestätigt die Vermutung, dass der verbleibende Regressor den Anteil der beiden ursprünglichen Regressoren übernommen hat. Die Quadratsumme der Residuen ist dabei nur unwesentlich von 2563.45 auf 2677.47 gestiegen. Gleichzeitig ist aber der absolute t -Wert des verbleibenden Regressors von 10.2 auf 49.02 gestiegen. Er hat also statistisch plötzlich viel mehr Bedeutung erhalten. Dies bestätigt, dass die t -Kenngrösse bei einer Kollinearität nichts über die Bedeutung eines Regressors im Gesamtmodell aussagt. Der berechnete Wert gibt eben nur einen Hinweis auf die Zunahme der Quadratsumme für den nächsten Eliminationsschritt an. Diese Zunahme ist aber naturgemäss klein, da ja der verbleibende Regressor den ganzen Einfluss der beiden früheren Regressoren übernehmen kann.

Als Nebenbemerkung kann festgestellt werden, dass die Multikollinearität in diesem Beispiel keinen negativen Einfluss auf die Prognose hat. Dies gilt, solange die starke Korrelation zwischen den beiden Temperaturen auch in der Zukunft erhalten bleibt.

Tabelle 6-12: Modell mit Betontemperaturen ohne T_H15_DO

Regressor	Coefficient	Std Err	t-stat	p-value	VIF	Std Coeff
Constant	29.3284	3.716	7.893	5.766E-15	0	0
T1(W_LEVEL)	25.786	0.1761	146.5	0	5.247	0.959271
T2(W_LEVEL)	7.18701	0.1108	64.88	0	3.267	0.335335
T3(W_LEVEL)	2.37925	0.0869	27.38	2.174E-133	2.271	0.117967
T4(W_LEVEL)	0.276952	0.07352	3.767	0.0001718	2.128	0.0157138
T_H12_UP	-0.408138	0.04302	-9.486	9.446E-21	5.387	-0.0629506
T_H12_MI	0.107371	0.1132	0.9485	0.343	10	0.00857678
T_H12_DO	-1.63306	0.03331	-49.02	5.051E-310	13.12	-0.50761
T_H15_UP	-1.4262	0.09094	-15.68	2.527E-51	9.374	-0.137294
T_H15_MI	6.01232	0.6882	8.736	6.555E-18	6.207	0.0622282
exp(-t)	0.565916	0.2525	2.241	0.02518	3.257	0.0115636
SS Res	MS Res	R2	F	signif F	DW	Observ
2677.47	1.84653	0.98815	1.209E4	0	0.08	1461
Regression: 01/01/1992-31/12/1995			Prediction: 01/01/1996-31/12/1998			
SS Res	MS' Res		SS Res	MS' Res		
2677.47	1.83263		4261.08	3.88785		

6.4.2 Beseitigung der Multikollinearität durch Principal-Component-Regression

Die zweite Strategie zur Beseitigung der Kollinearität ist die Principal-Component-Regression. Die Eigenwerte sind in Tabelle 6-13 zusammengestellt. Wie man sieht, ist der letzte Eigenwert (in der Tabelle ist der Singulärwert $\sigma = \sqrt{\lambda}$ aufgeführt) sehr klein. Durch sein Weglassen verschwindet die Kollinearität. Auch die entsprechende t -Testgrösse ist klein, sodass man nur einen geringen Zuwachs in den Residuen erwarten muss. Das Weglassen des Eigenvektors mit dem kleinsten Eigenwert führt zum Modell in Tabelle 6-14. Tatsächlich sind die hohen VIF-Werte verschwunden. Die

Regressionskoeffizienten haben sich gegenüber dem ursprünglichen Modell nur wenig verändert und Quadratsumme der Residuen ist nur wenig angewachsen. Die beiden Temperaturen T_H12_DO und T_H15_DO haben nun hohe absolute t -Werte (46.3 und 55.84), vergleichbar mit demjenigen, wenn eine der beiden Temperaturen eliminiert wurde (49.02).

Tabelle 6-13: Eigenwerte für Modell mit Betontemperaturen

Eigenvektor	V1	V2	V3	V4	V5	
sigma	2.09551	1.69661	1.13919	1.01823	0.837929	
t-stat	-114.819	255.924	-88.5631	122.868	115.184	
Eigenvektor	V6	V7	V8	V9	V10	V11
sigma	0.455223	0.429993	0.373433	0.344746	0.196466	0.0683848
t-stat	-51.9707	4.89982	-25.5971	87.5397	-17.1298	-0.733873

Tabelle 6-14: Modell mit Betontemperaturen, ein Eigenvektor weggelassen

Regressor	Coefficient	StdErr	t-stat	p-value	VIF	Std Coeff
Constant	34.3062	3.677	9.33	3.822E-20	0	0
T1(W_LEVEL)	25.9548	0.1731	150	0	5.292	0.965553
T2(W_LEVEL)	7.13354	0.1082	65.95	0	3.251	0.332841
T3(W_LEVEL)	2.24923	0.08555	26.29	6.866E-125	2.296	0.11152
T4(W_LEVEL)	0.304555	0.07188	4.237	2.406E-5	2.123	0.0172799
T_H12_UP	-0.434756	0.04176	-10.41	1.565E-24	5.294	-0.067056
T_H12_MI	-0.0324894	0.1119	-0.2904	0.7715	10.19	-0.00259526
T_H12_DO	-0.87482	0.01889	-46.3	4.898E-288	4.401	-0.271924
T_H15_UP	-1.05036	0.09208	-11.41	6.332E-29	10.03	-0.101114
T_H15_MI	4.87707	0.6832	7.139	1.485E-12	6.381	0.0504783
T_H15_DO	-0.942027	0.01687	-55.84	0	2.806	-0.261832
exp(-t)	0.0115119	0.2525	0.0456	0.9636	3.396	0.000235227
SS Res	MS Res	R2	F	signif F	DW	Observ
2564.41	1.76978	0.98865	1.147E4	0	0.07	1461
Regression: 01/01/1992-31/12/1995			Prediction: 01/01/1996-31/12/1998			
SS Res	MS' Res	SS Res		MS' Res		
2564.41	1.75524	3621.02		3.30385		

Abschliessend wird für diesen einfachen Fall noch gezeigt, wie man aus dem weggelassenen Eigenvektor eine Beziehung zwischen den beiden korrelierten Temperaturen finden kann. Der Eigenvektor ist

$$V_{11} = 0.011 \quad -0.011 \quad -0.015 \quad 0.007 \quad -0.023 \quad -0.013 \quad 0.685 \quad 0.086 \quad -0.019 \quad -0.722 \quad -0.020$$

Wie in Kapitel 4 beschrieben wurde, bedeutet das Weglassen des Eigenvektors 11 (nur die grössten Anteile berücksichtigt)

$$\alpha_{11} \approx 0.685b_7 - 0.722b_{10} = 0. \quad (6.1)$$

Daraus erhält man eine Zusatzbedingung für die Regressionskoeffizienten, nämlich

$$b_{10} = (0.685/0.722)b_7 = 0.95b_7. \quad (6.2)$$

b_7 ist der Koeffizient von T_H12_DO und b_{10} derjenige von T_H15_DO. Die obige Beziehung ist für die standardisierten Regressionskoeffizienten in Tabelle 6-14 ungefähr eingehalten ($b_7 = -0.271924$ und $b_{10} = -0.261832$). Für die standardisierten Temperaturen T_H12_DO und T_H15_DO bedeutet das

$$b_7 \cdot T_{H12_DO} + b_{10} \cdot T_{H15_DO} = b_7 (T_{H12_DO} + 0.95 \cdot T_{H15_DO}) \quad (6.3)$$

Man kann also die beiden Temperaturen zu einem einzigen Regressor zusammenfassen. Man beachte, dass der Faktor 0.95 für die standardisierten Temperaturen gilt. Für die ursprünglichen Temperaturen muss der Faktor noch umgerechnet werden. Dies ist in den Regressionskoeffizienten in Tabelle 6-14 bereits berücksichtigt. Diese können direkt benützt werden, um einen neuen Regressor zu definieren:

$$x_{\text{neu}} = -0.87482 \cdot T_{H12_DO} - 0.942027 \cdot T_{H15_DO} \quad (6.4)$$

Verwendet man diesen neuen Regressor anstelle der beiden ursprünglichen Temperaturen, so ist keine Multikollinearität mehr vorhanden.

Diese Methode zur Bildung neuer Regressoren wird natürlich sofort viel komplizierter, wenn mehrere Regressoren an einem Eigenvektor beteiligt sind und wenn mehrere Eigenvektoren weggelassen werden. Ein ähnliches Beispiel ist in Chatterjee (1977, S.180) gegeben.

Wenn das Problem der Kollinearität gelöst ist, sei es durch Weglassen eines Regressors oder mit Hilfe der Principal-Component-Regression, können die nicht signifikanten Regressoren eliminiert werden. Wenn man von der Principal-Component-Lösung (Tabelle 6-14) ausgeht, können 2 Regressoren eliminiert werden. Der erste ist die Exponentialfunktion, der zweite die Temperatur T_H12_MI. Die Resultate bleiben nach dieser Elimination im Wesentlichen unverändert, wie die Zusammenfassung in Tabelle 6-15 zeigt. Die Residuen und das Korrelogramm sind in Bild 6-15 und Bild 6-16 festgehalten.

Tabelle 6-15: Modell mit Betontemperaturen, 9 Regressoren, ein Eigenvektor weggelassen

Regression: 01/01/1992-31/12/1995		Prediction: 01/01/1996-31/12/1998	
SS Res	MS' Res	SS Res	MS' Res
2565.54	1.75602	3636.72	3.31818

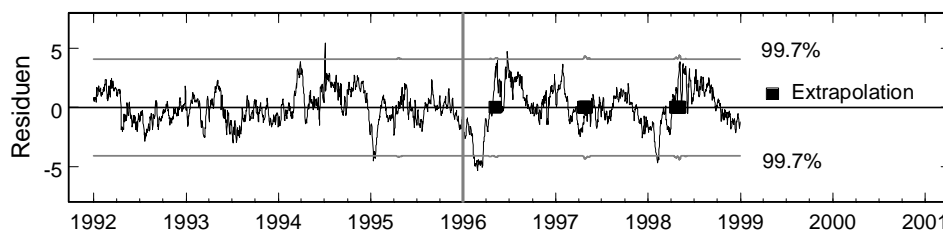


Bild 6-15: Residuen, Modell mit Betontemperaturen, 9 Regressoren, ein Eigenvektor weggelassen

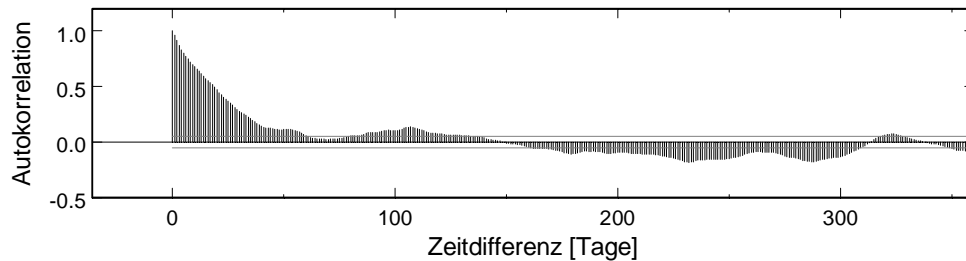


Bild 6-16: Korrelogramm, Modell mit Betontemperaturen, 9 Regressoren, ein Eigenvektor weglassen

6.4.3 Autokorrelation der Fehler

Der zweite Weg, der vom theoretischen Standpunkt aus vorzuziehen ist, ist die Berücksichtigung der Autokorrelation der Fehler. Die vorgeschlagene Methode mit der iterativen Berechnung von ρ führt zunächst nicht zum Erfolg. Für den Autokorrelationsparameter $\rho = 1$ bleiben sinusförmige Residuen mit einer grossen Amplitude (± 8 mm). Variiert man ρ zwischen 0 und 1, so entdeckt man, dass die sinusförmigen Residuen erst in der Nähe von 1 auftauchen. Andererseits ist der Effekt von ρ kaum sichtbar zwischen 0 und etwa 0.8. Versucht man den Wert von Hand zu optimieren, findet man etwa $\rho = 0.9$. Wenn man noch die beiden nichtsignifikanten Regressoren (T_H12_MI und T_H15_MI) eliminiert, erhält man das Modell in Tabelle 6-16. Sowohl die Regressionskoeffizienten wie auch die t -Kenngrössen sind mit dem transformierten Problem (verallgemeinerte Differenzenbildung) berechnet, gelten aber auch für das ursprüngliche Problem. Der grösste VIF-Wert ist um 14, was akzeptabel ist. Die zusammenfassenden Grössen wie die Quadratsummen und das Bestimmtheitsmass gelten nur für das transformierte Problem. Der Durbin-Watson-Kennwert ist wesentlich grösser als bei der klassischen Methode. Um den Vergleich mit andern Modellen zu ermöglichen, sind auf der untersten Zeile die Quadratsummen der Residuen für das ursprüngliche Modell gegeben. Die Residuen in der Prognoseperiode sind besser als diejenigen der klassischen Methode, wie der Vergleich in Tabelle 6-16 zeigt. (Siehe auch Zusammenstellung der Residuenplots am Ende dieses Kapitels.) Die Residuen sind in Bild 6-17 geplottet. Das Korrelogramm in Bild 6-18 zeigt, dass noch eine leichte Autokorrelation vorhanden ist, die aber akzeptiert werden kann.

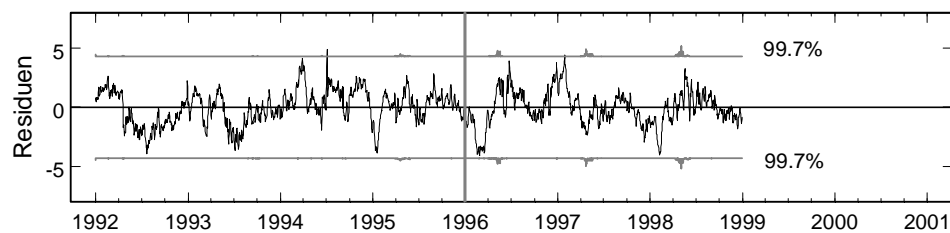
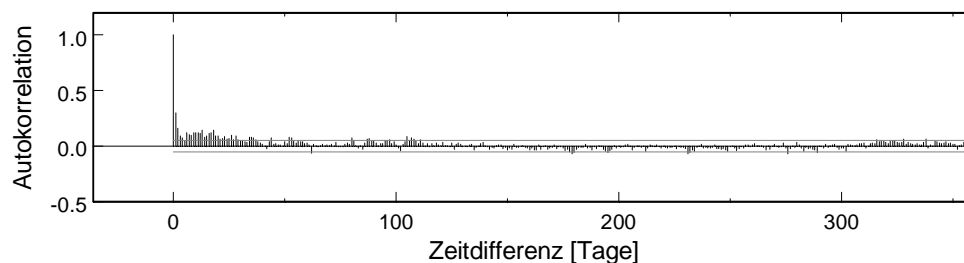
Das hier beschriebene Vorgehen ist natürlich nicht gedacht für eine generelle Anwendung, aber es zeigt, dass die verallgemeinerte Methode der kleinsten Quadrate funktioniert, wenn der Parameter ρ richtig berechnet wird. Das Problem ist ja, dass es sich hier um eine nichtlineare Regression handelt, bei der die Regressionsparameter und der Autokorrelationsparameter ρ gleichzeitig optimiert werden müssen. Die iterative Lösung, die hier verwendet wurde, ist nur die einfachste Variante von den vielen, die in der Literatur zu finden sind. Das zweite Problem, das hier möglicherweise eine Rolle spielt, ist, dass die Autokorrelation der Fehler auch höherer Ordnung sein kann. Die Methode müsste dann entsprechend angepasst werden. Sowohl die numerische Behandlung der nichtlinearen Regression, wie auch die Behandlung anderer Autokorrelationsstrukturen liegen aber ausserhalb des Rahmens dieser Studie.

Tabelle 6-16: Modell mit Betontemperaturen, $\rho = 0.9$ (Optimierung von Hand)

Regressor	Coefficient	Std Err	t-stat	p-value	VIF	Std Coeff
Constant	57.0584	0.7021	81.26	0	0	0
T1(W_LEVEL)	25.4259	0.2794	91.01	0	2.198	0.845923
T2(W_LEVEL)	7.33344	0.1851	39.62	2.353E-233	1.743	0.327928
T3(W_LEVEL)	1.92538	0.1367	14.08	2.671E-42	1.296	0.100493
T4(W_LEVEL)	0.678253	0.1115	6.082	1.513E-9	1.282	0.0431809
T_H12_UP	-0.393797	0.07179	-5.485	4.868E-8	2.961	-0.0591799
T_H12_DO	-0.776487	0.08113	-9.571	4.372E-21	13.2	-0.218055
T_H15_UP	-0.716141	0.1174	-6.098	1.373E-9	3.014	-0.0663771
T_H15_DO	-1.11142	0.0956	-11.63	6.176E-30	13.78	-0.27058
SS Res	MS Res	R2	F	signif F	DW	Observ
181.485	0.12499	0.94292	2987	0	1.40	1461
Regression: 01/01/1992-31/12/1995			Prediction: 01/01/1996-31/12/1998			
SS Res	MS' Res		SS Res	MS' Res		
3027.93	2.0725		2337.15	2.13243		

Tabelle 6-17: Vergleich klassische Methode und verallgemeinerte Methode (Tabelle 6-15 und Tabelle 6-16)

		Regression: 01/01/1993-31/12/1995		Prediction: 31/12/1994-31/12/1998	
	Reg	SS Res	MS' Res	SS Res	MS' Res
$\rho = 0$	9	2565.54	1.75602	3636.72	3.31818
$\rho = 0.9$	8	3027.93	2.0725	2337.15	2.13243

**Bild 6-17: Residuen, Modell mit Betontemperaturen, $\rho = 0.9$ (Optimierung von Hand)****Bild 6-18: Korrelogramm, Modell mit Betontemperaturen, $\rho = 0.9$ (Optimierung von Hand)**

6.5 Modell mit Wärmeleitung

Der Einbezug der Betontemperaturen hat die Regression gegenüber dem klassischen Ansatz mit Sinus und Kosinus nur bedingt verbessert. Eine mögliche Erklärung ist, dass die Temperaturverteilung in der Mauer mit den Messungen am Rand und in der Mitte zu wenig gut erfasst wird. Es handelt sich hier um ein Wärmeleitungsproblem, wie es in Kapitel 5 beschrieben wurde. Die Innentemperaturen treten gegenüber den Aussentemperaturen verkleinert und zeitlich verzögert auf. Das Phänomen ist gut sichtbar aus den Temperaturverläufen in Bild 6-4.

Die theoretischen Aspekte der eindimensionalen Wärmeleitung wurden bereits in Kapitel 5 beschrieben. Als erstes soll hier versucht werden, die Betontemperatur im Innern der Mauer auf Messhorizont MH12 aus den Betontemperaturen an der Oberfläche zu berechnen. Da nur die Lösungen für ein halbumendliches Gebiet verwendet werden, kann der Einfluss von beiden Seiten annäherungsweise durch eine Überlagerung beider Einflüsse berechnet werden. Dies ist mathematisch nicht ganz korrekt, da die Temperatur am einen Rand den andern Rand nicht beeinflussen sollte. Die Messpunkte an der Luftseite und der Wasserseite sind nicht ganz an der Oberfläche, sondern 1.5 m im Innern, sodass die Distanz zur Mitte als 6 m angenommen werden kann. Für die Wärmediffusion wurde der Wert $0.1 \text{ m}^2/\text{Tag}$ eingesetzt. Bild 6-19 zeigt den Vergleich mit der gemessenen Temperatur (auf Mittelwert zentriert). Die Übereinstimmung ist erstaunlich gut, wenn man die Einfachheit des Modells berücksichtigt. (Der treppenförmige Verlauf der gemessenen Temperatur ist eine Folge der kleinen Auflösung von nur zwei Ziffern. Die markante Spitze Mitte 1994 ist wahrscheinlich ein Messfehler.)

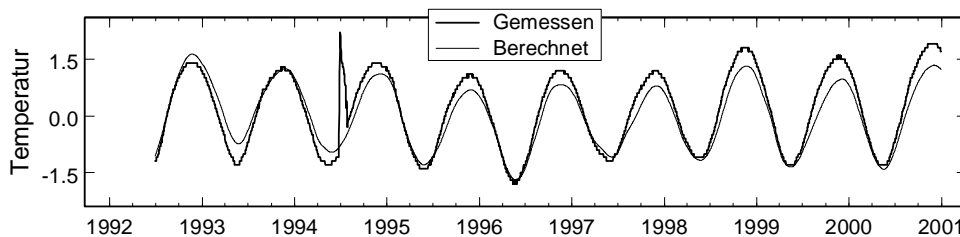


Bild 6-19: Berechneter Temperaturverlauf in der Mitte der Mauer

In den kommenden Modellen werden für die Temperaturberechnung jeweils nur die Einflüsse vom näher gelegenen Rand berücksichtigt.

6.5.1 Klassische Regression und Ridge-Regression

Um die Berechnung einigermaßen übersichtlich zu halten und nicht zu viele Regressoren und Multikollinearitäten zu erhalten, kann man sich auf die Temperaturen in 2 und in 4 Tiefe beschränken. Damit hat man neben den 6 gemessenen noch weitere 8 berechnete Temperaturen im Modell. Da die berechneten Temperaturen von vergangenen Daten abhängen, kann man die Berechnung erst ab einem Zeitpunkt durchführen, bei dem genügend Werte aus der Vergangenheit zur Verfügung stehen. Hier wurde der Beginn der Regressionsperiode 180 Tage nach Beginn der Messperiode festgelegt.

Damit ist die Regressionsperiode nicht mehr 4, sondern nur noch $3\frac{1}{2}$ Jahre lang. Es ergibt sich eine gewisse Schwierigkeit beim Vergleich der Modelle, da wie früher festgestellt wurde, verschiedene Regressionsperioden zu verschiedenen guten Resultaten führen können.

Das Modell mit 4 Polynomen, 6 gemessenen und 8 berechneten Temperaturen, insgesamt also 18 Regressoren, zeigt bei 13 Regressoren VIF-Werte über 10, die grössten über 1000. Um die Anzahl der Regressoren zu reduzieren, kann man wieder verschiedene Strategien anwenden. Die eine ist die, einfach Regressoren zu eliminieren bis keine Multikollinearität mehr vorhanden sind. Dies führt zum Modell mit 5 Regressoren in Tabelle 6-18. temp(T_H12_DO|4) bezeichnet die die aus T_H12_DO abgeleitete Temperatur in 4 m Tiefe. Interessant ist, dass die Abweichungen kleiner sind als beim Modell mit 5 gemessenen Temperaturen (Tabelle 6-15), obwohl hier nur Messwerte einer einzigen Temperatur verwendet wurden.

Die Ridge-Regression führt je nach Verzerrungsparameter k zu verschiedenen Modellen, wie die Zusammenstellung in Tabelle 6-19 zeigt. Für jeden Fall wurden die Regressoren schrittweise eliminiert, bis all VIF-Werte unter 10 waren. Das Beispiel zeigt schön, wie man mit einer Variation des Verzerrungsparameters das Modell optimieren kann.

Tabelle 6-18: Modell mit Wärmeleitung

Regressor	Coefficient	Std Err	t-stat	p-value	VIF	Std Coeff
Constant	51.4513	0.08108	634.6	0	0	0
T1(W_LEVEL)	26.6533	0.1585	168.1	0	3.904	1.01228
T2(W_LEVEL)	7.76341	0.09571	81.12	0	2.346	0.378663
T3(W_LEVEL)	2.01985	0.0754	26.79	8.758E-126	1.498	0.0999171
T_H12_DO	-2.02508	0.01077	-188	0	1.258	-0.642689
temp(T_H12_DO 4)	-1.88981	0.07463	-25.32	5.993E-115	4.626	-0.165992
SS Res	MS Res	R2	F	signif F	DW	Observ
2222.85	1.74341	0.98816	2.128E4	0	0.08	1281
Regression: 29/06/1992-31/12/1995			Prediction: 01/01/1996-31/12/1998			
SS Res	MS' Res	SS Res		MS' Res		
2222.85	1.73525	3002.23		2.73927		

Tabelle 6-19 Modelle mit Wärmeleitung, schrittweise Elimination mit Ridge-Regression

		Regression: 29/06/1992-31/12/1995		Prediction: 01/01/1996-31/12/1998	
	Reg	SS Res	MS' Res	SS Res	MS' Res
k = 0	5	2222.85	1.73525	3002.23	2.73927
k = 0.001	5	2224.12	1.73624	3009.33	2.74574
k = 0.002	6	1542.71	1.2043	2523.28	2.30226
k = 0.005	6	1556.52	1.21508	2491.25	2.27304
k = 0.01	6	1602.94	1.25132	2475.31	2.25849
k = 0.02	7	1773.7	1.38462	3347.72	3.05449

6.5.2 Principal-Component-Regression

Wie in Kapitel 4 anhand eines numerischen Beispiels gezeigt wurde, bildet die Principal-Component-Regression aus zwei ähnlichen Regressoren eine Kombination. Die abhängige Variable, die approximiert werden soll, wird dabei nicht berücksichtigt. In diesem Sinne kann die Principal-Component-Regression nicht den besser passenden Regressor auswählen, das macht die klassische Regression. Auf der andern Seite vermeidet die Principal-Component-Regression, dass mehrere Regressoren, die nichts mit der abhängigen Variablen zu tun haben, trotzdem so kombiniert werden, dass eine Approximation entsteht (typischerweise über grosse Koeffizienten mit verschiedenen Vorzeichen). Aus der ersten Überlegung wird klar, dass mit der Principal-Component-Regression oft nicht das optimale Modell gefunden wird, besonders in den Fällen, wo man aus mehreren ähnlichen Regressoren auswählen muss, wie in diesem Fall mit den berechneten Temperaturen.

Es kann hier keine einfache Strategie angegeben werden, wie man die Variablenauswahl mit der Principal-Component-Regression durchführen soll. Da man in jedem Schritt sowohl Eigenvektoren weglassen wie auch Regressoren eliminieren kann, gibt es viele Vorgehensmöglichkeiten. Mehrere davon wurden ausprobiert, aber keine hat sich als allgemein beste erwiesen. Es ist fast immer nötig, einige Varianten auszuprobieren, bis man sich ein Bild machen kann.

Das Ausgangsmodell ist wieder das Modell mit 4 Tschebyscheff-Polynomen, 6 gemessenen und 8 berechneten Temperaturen. Um die Multikollinearität zu beseitigen, muss man etwa 8 Eigenvektoren weglassen. Ein gutes Bild geben die Eigenwerte und die dazugehörigen t -Kenngrössen, die in Tabelle 6-20 aufgelistet sind. Da die Singulärwertzerlegung benützt wurde, sind anstatt der Eigenwerte die Singulärwerte $\sigma = \sqrt{\lambda}$ (sigma) gezeigt. Rein von den Eigenwerten her kann man nicht entscheiden, wo man abschneiden soll, da die Werte sich ziemlich kontinuierlich ändern, ohne klare Sprünge. Bei den t -Kenngrössen, vermutet man einen grossen Sprung in der Quadratsumme der Residuen, von 12 auf 11 Eigenvektoren. Tatsächlich findet der Sprung von 11 auf 10 Eigenvektoren statt, wie man feststellt, wenn man die Berechnung durchführt. Das ist auch der Punkt, wo die VIF-Werte unter 10 fallen.

Tabelle 6-20: Singulärwerte für Modell mit Wärmeleitung

	V1	V2	V3	V4	V5	V6	V7	V8	V9
Sigma	3.11548	2.16845	1.10739	0.933764	0.793533	0.529353	0.449395	0.394638	0.297116
t-stat	192.323	-221.601	-46.7018	-112.425	125.841	-80.1263	37.0983	-81.1442	34.8156

	V10	V11	V12	V13	V14	V15	V16	V17	V18
Sigma	0.25218	0.19906	0.12154	0.10278	0.06613	0.043699	0.0368386	0.031658	0.01708
t-stat	-44.7478	17.2026	-5.48059	6.89917	-2.3466	-7.12465	-1.55732	-0.462134	14.8993

Als ersten Schritt reduziert man also die Zahl der Eigenvektoren auf 11. Von den Regressoren sind dann einige nicht signifikant. Die 18 Regressoren kann man deshalb schrittweise auf 13 reduzieren. Dadurch bekommt man aber wieder erneut eine Multikollinearität, wie der höchste VIF-Wert von 50 angibt. Nachdem man 2 weitere Eigen-

vektoren weggelassen hat, hat man also jetzt 13 Regressoren und 9 Eigenvektoren. Bei diesem Modell sind alle Regressoren signifikant und die VIF-Werte alle unter 10. Die Quadratsummen sind $SS_{\text{Res}}=1.06$ und $MS_{\text{Res}}= 2.23$. Man kann aber das Modell noch weiter reduzieren. Bei der Elimination von weiteren Regressoren können immer wieder neue Multikollinearitäten auftauchen, sodass man parallel dazu auch die Zahl der Eigenvektoren laufend reduzieren muss. Ein Modell das nach diesem Vorgehen gefunden werden kann, ist in Tabelle 6-21 angegeben. Von den Residuen her ist dieses Modell gleichwertig mit den früheren, die mit der Ridge-Regression gefunden wurden. Die Residuen sind in Bild 6-20 geplottet.

Tabelle 6-21: Modell mit Wärmeleitung, Principal-Component-Regression, 1 Eigenvektor weggelassen

Regressor	Coefficient	Std Err	t-stat	p-value	VIF	Std Coeff
Constant	49.5932	0.1273	389.7	0	0	0
T1(W_LEVEL)	26.4324	0.1448	182.5	0	3.475	1.00389
T2(W_LEVEL)	7.51159	0.08997	83.49	0	2.212	0.36638
T3(W_LEVEL)	2.11668	0.07234	29.26	2.134E-144	1.471	0.104707
T_H12_DO	-0.73701	0.01499	-49.16	1.539E-296	2.601	-0.233901
T_H15_DO	-0.993804	0.007721	-128.7	0	0.5535	-0.282516
temp(T_H12_DO 2)	-1.29132	0.05732	-22.53	7.031E-95	10.8	-0.218463
SS Res	MS Res	R2	F	signif F	DW	Observ
2081.78	1.63405	0.98891	1.893E4	0	0.07	1281
Regression: 29/06/1992-31/12/1995			Prediction: 01/01/1996-31/12/1998			
SS Res	MS' Res	SS Res		MS' Res		
2081.78	1.62512	2414.22		2.20276		

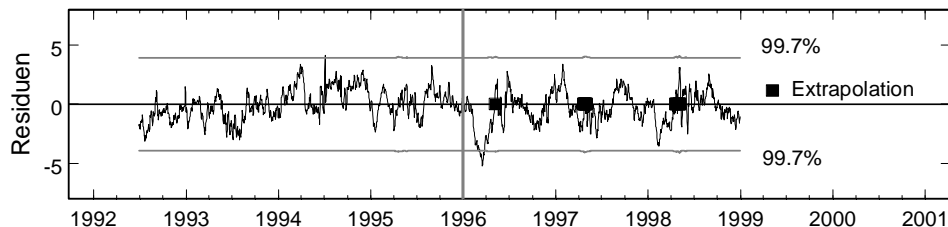


Bild 6-20: Residuen, Modell mit Wärmeleitung, Principal-Component-Regression, 1 Eigenvektor weggelassen

6.5.3 Autokorrelation der Fehler

Wie der Durbin-Watson-Test im letzten Modell deutlich zeigt, sind die Fehler autokorreliert. Eine Berechnung mit der verallgemeinerten Methode der kleinsten Quadrate bringt aber zunächst wieder nicht die gewünschte Lösung. Die Residuen enthalten eine Langzeitkomponente, die das Resultat unbrauchbar macht. Auch hier kann aber durch eine manuelle Anpassung des Autokorrelationsparameters ρ eine gute Anpassung erreicht werden. Ob das Problem am Algorithmus für die Berechnung von ρ liegt, oder daran, dass die Verteilung der Fehler komplexer ist als angenommen, kann an dieser Stelle nicht beantwortet werden. Es müssten zusätzliche Untersuchungen gemacht werden, die weit über den Rahmen dieser Studie hinausgehen würden.

Das Ausgangsmodell ist wieder jenes mit 4 Tschebyscheff-Polynomen, 6 gemessenen und 8 berechneten Temperaturen. Mit $\rho = 0.92$ können die Regressoren schrittweise eliminiert werden bis alle p -Werte unter 1% sind. Das führt zu einem Modell mit 9 Regressoren. Vier weitere Regressoren müssen eliminiert werden, um die VIF-Werte zu reduzieren. Wenn wieder zwei Regressoren eingefügt werden, erhält man das Modell in Tabelle 6-22. Die ganze Variablenauswahl ist ein Ausprobieren und hängt auch stark von der Wahl von ρ ab. Die schrittweise Elimination führt nicht direkt zum Ziel, da eine starke Multikollinearitäten vorhanden sind. Beim vollen Modell ist der maximale VIF-Wert bei 500.

Die Quadratsummen der Residuen sind deutlich besser als jene der klassischen Regression, allerdings mit 7 gegenüber 6 Regressoren (5 Eigenwerten). Der Vergleich dieser beiden Modelle mit dem Modell der klassischen Methode ist in Tabelle 6-23 zu sehen. (Siehe auch Zusammenstellung der Residuenplots am Ende dieses Kapitels.) Die Residuen sind in Bild 6-21 geplottet.

Tabelle 6-22: Modell mit Wärmeleitung, $\rho = 0.92$ (Optimierung von Hand)

Regressor	Coefficient	Std Err	t-stat	p-value	VIF	Std Coeff
Constant	46.063	0.2894	159.2	0	0	0
T1(W_LEVEL)	26.6963	0.3588	74.41	0	2.579	0.938418
T2(W_LEVEL)	7.69873	0.1935	39.79	1.202E-225	1.562	0.390614
T3(W_LEVEL)	1.82276	0.1418	12.85	1.23E-35	1.159	0.108687
T_H12_DO	-0.82139	0.05263	-15.61	2.232E-50	4.404	-0.257242
temp(T_H12_UP 2)	-1.20187	0.2343	-5.13	3.349E-7	4.977	-0.0898782
temp(T_H15_UP 2)	2.9879	0.3197	9.346	3.932E-20	3.858	0.144186
temp(T_H15_DO 2)	-2.89414	0.194	-14.92	1.686E-46	11.02	-0.388874
SS Res	MS Res	R ²	F	signif F	DW	Observ
128.144	0.100663	0.92148	2150	0	1.70	1281
Regression: 29/06/1992-31/12/1995			Prediction: 01/01/1996-31/12/1998			
SS Res	MS' Res		SS Res	MS' Res		
1613.36	1.25945		2021.44	1.84438		

Tabelle 6-23: Vergleich klassische und verallgemeinerte Methode (Tabelle 6-21 und Tabelle 6-22)

		Regression: 29/06/1992-31/12/1995		Prediction: 31/12/1994-31/12/1998	
	Reg	SS Res	MS' Res	SS Res	MS' Res
$\rho = 0$	6	2081.78	1.62512	2414.22	2.20276
$\rho = 0.92$	7	1613.36	1.25945	2021.44	1.84438

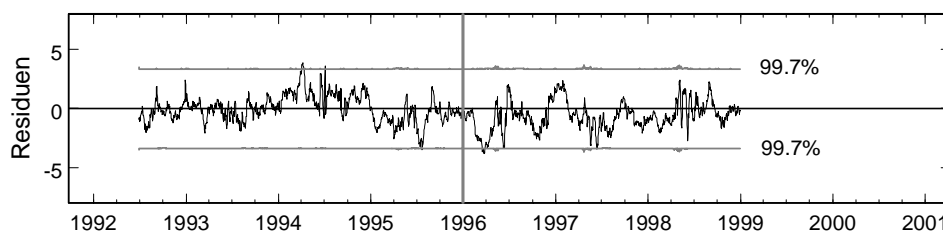


Bild 6-21: Residuen, Modell mit gemessenen und berechneten Betontemperaturen, 7 Regressoren, $\rho = 0.92$

6.6 Weitere Modelle

6.6.1 Modell mit Kriechen

Regressoren für Kriechverformungen können aus den Regressoren für die elastische Verformung abgeleitet werden, wie dies in Kapitel 5 gezeigt wurde. Für die vier Tschebyscheff-Polynome werden die zugehörigen Kriech-Regressoren zum Modell mit den 6 gemessenen und den 8 berechneten Betontemperaturen hinzugefügt. Wieder wurden für verschiedene Verzerrungsparameter die Regressoren eliminiert bis alle VIF-Werte unter 10 waren. Die Resultate sind in Tabelle 6-24 zusammengestellt. In der Spalte "Reg" ist die Anzahl der Regressoren und in der Spalte "Creep" sind die beteiligten Kriechfunktionen angegeben. In den meisten Modellen ist die Kriechfunktion nur in der ersten Potenz enthalten (T1), in einigen überhaupt nicht. Betrachtet man nur die Regressionsperiode, so können mit dem Kriechansatz die Residuen halbiert werden. In der Prognoseperiode jedoch sind die Modelle ohne Kriechen deutlich besser als jene mit Kriechen. Auch die Möglichkeit, dass mit dem Kriechansatz die Autokorrelation der Fehler aufgehoben werden könnte, hat sich nicht bestätigt. Die Durbin-Watson-Kennwerte waren in allen Fällen um $DW = 0.10$.

Tabelle 6-24: Modelle mit gemessenen und berechneten Betontemperaturen und Kriechen, schrittweise Elimination mit Ridge-Regression mit verschiedenen Verzerrungsparametern

			Regression: 29/06/1992-31/12/1995		Prediction: 01/01/1996-31/12/1998	
	Reg	Creep	SS Res	MS' Res	SS Res	MS' Res
k = 0	5		2222.85	1.73525	3002.23	2.73927
k = 0.0005	7	T1	1046.07	0.816604	6493.34	5.92458
k = 0.001	5		1914.79	1.49476	2497.92	2.27912
k = 0.002	6	T1	1608.62	1.25575	8248.39	7.52591
k = 0.005	6	T1	1641.29	1.28126	8155.87	7.44149
k = 0.01	6	T1	1967.56	1.53596	5609.53	5.11819
K = 0.02	8	T1, T2	1477.04	1.15303	4754.52	4.33807

6.6.2 Modell mit Lufttemperatur

Ein weiteres Modell berücksichtigt den Temperatureinfluss über die Lufttemperatur. Die Lufttemperatur ist nur eine indirekte Einflussgrösse, da die Verformungen direkt von der Temperaturverteilung in der Mauer abhängen. Da die Temperaturen in der Mauer viel weniger rasch schwanken als die Lufttemperatur, ist es zweckmässig zeitliche Mittelwerte der Lufttemperatur als Einflussvariablen zu wählen. Das Modell beinhaltet 4 Tschebyscheff-Polynome und die Lufttemperatur, sowie die Mittelwerte über 7, 14, 30, 60, 120 und 180 Tage. Um das Mittel von 180 berechnen zu können, beginnt die Regressionsperiode mit dem 181. Tag.

Der maximale VIF-Wert ist etwa 80. Die Elimination der Regressoren erfolgt mit der klassischen Regression schrittweise bis alle VIF-Werte unter 10 sind. Die Ridge-Regression bringt bezüglich der Variablenauswahl keine Änderung. Tabelle 6-25 zeigt das Modell. Die Regressionsqualität ist schlechter als beim Modell mit Sinus und Kosinus, obwohl mehr Information verwendet wurde. Daher ist das Modell mit der Lufttemperatur eher ungeeignet.

Tabelle 6-25: Modell mit Lufttemperatur

Regressor	Coefficient	Std Err	t-stat	p-value	VIF	Std Coeff
Constant	46.935	0.08143	576.4	0	0	0
T1(W_LEVEL)	26.474	0.1558	170	0	2.446	1.00547
T2(W_LEVEL)	6.78669	0.1072	63.32	0	1.91	0.331023
T3(W_LEVEL)	1.94256	0.09157	21.21	8.801E-86	1.434	0.096094
T_AIR_M 7	-0.197277	0.01326	-14.87	2.939E-46	3.596	-0.106684
T_AIR_M 60	-1.3061	0.01886	-69.25	0	5.288	-0.602313
SS Res	MS Res	R2	F	signif F	DW	Observ
3424.45	2.68584	0.98176	1.372E4	0	0.04	1281
Regression: 29/06/1992-31/12/1995			Prediction: 01/01/1996-31/12/1998			
SS Res	MS' Res	SS Res		MS' Res		
3424.45	2.67326	4405.06		4.01921		

6.7 Zusammenfassung Schlegeis

Das klassische Modell mit Sinus- und Kosinus-Termen für die saisonalen Einflüsse funktioniert gut mit der Methode der kleinsten Quadrate. Die Elimination der Regressoren kann nach den p -Werten erfolgen, aber einige zusätzliche Regressoren können ebenfalls eliminiert werden, obwohl sie gemäss ihrem p -Wert statistisch signifikant wären. Dieses Problem kommt daher, dass die Autokorrelation der Fehler vernachlässigt wurde. Die Autokorrelation der Fehler ist klar vorhanden, wie der Durbin-Watson-Kennwert und das Korrelogramm zeigen. Diese kann mit der verallgemeinerten Methode der kleinsten Quadrate berücksichtigt werden. Die Variablenauswahl kann dann rein anhand der p -Werte erfolgen. Theoretisch sollte die Methode auch effizientere Regressionskoeffizienten liefern, was aber bei diesem Modell nicht bestätigt wurde.

Benützt man anstatt den saisonalen Funktionen die Betontemperaturen, stösst man zusätzlich zur Autokorrelation der Fehler auf das Problem der Multikollinearität. Zwei Temperaturen an der Luftseite sind stark korreliert. Man kann die Kollinearität beseitigen indem entweder eine von beiden Temperaturen eliminiert wird oder indem in der Principal-Component-Regression ein Eigenvektor weggelassen wird. Die verallgemeinerte Methode der kleinsten Quadrate bringt beim Modell mit Betontemperaturen numerische Probleme bei der Bestimmung des Autokorrelationsparameters ρ . Um zu zeigen, dass die Methode grundsätzlich funktioniert, wurde der Parameter ρ von Hand optimiert.

Als nächstes werden zusätzlich zu den gemessenen auch berechnete Temperaturen eingeführt. Die Berechnung erfolgt an Hand der eindimensionalen Wärmeleitung. Dieses Modell hat bereits 18 Regressoren und viele Multikollinearitäten, was die Variablenauswahl erschwert. Eine Strategie ist, Regressoren zu eliminieren bis alle VIF-Werte unter 10 sind. Die schrittweise Elimination führt aber oft zum falschen Modell, da die Regressoren stark korreliert sind. Ein nützliches Verfahren ist hier die Ridge-Regression. Berechnet man mehrere Varianten mit verschiedenen Verzerrungsparametern, so kommt man zu einem guten Modell. Die Alternative zur Ridge-Regression ist die Principal-Component-Regression. Da man sowohl Regressoren wie auch Eigenwerte eliminierten kann, ergeben sich viele Vorgehensweisen, die nicht alle gleich gut

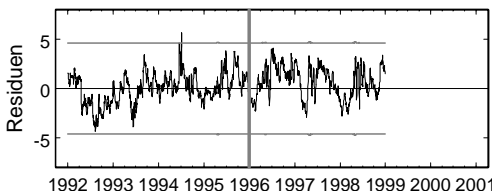
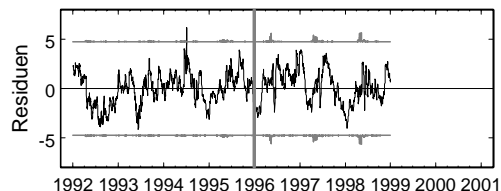
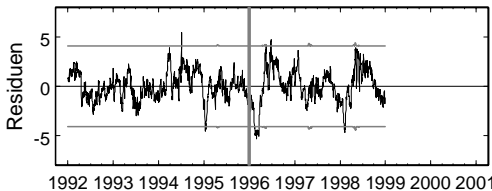
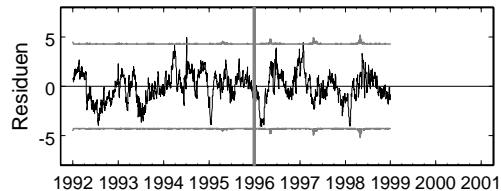
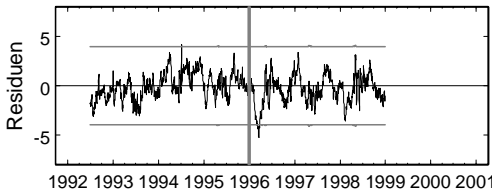
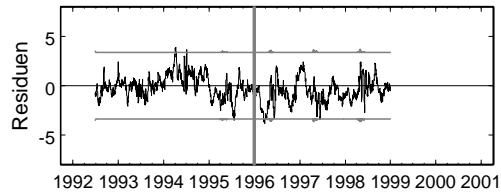
funktionieren. Bei der verallgemeinerten Methode musste wieder von Hand eingegriffen werden, um eine gute Regression zu bekommen.

Der Vollständigkeit halber wurden auch noch der Einfluss von Kriechen und der Lufttemperatur untersucht. Kriechen scheint wenig Einfluss zu haben und hilft auch nicht die Autokorrelation der Fehler zu verhindern. Die Lufttemperatur kann benützt werden, wenn Mittelwerte verwendet werden, aber das Modell war weniger gut als das Sinus-Kosinus-Modell.

Zum Schluss sind in Tabelle 6-26 die Residuen der wichtigsten Modelle nochmals zusammengestellt. Beim Vergleich stellt man zunächst fest, dass die Unterschiede zwischen den verschiedenen Modellen und den verschiedenen Methoden nicht so unterschiedlich sind, wie man es vielleicht erwartet hätte. Immerhin stellt man aber bei genauerem Hinsehen fest, dass vor allem der Einbezug der Wärmeleitung eine Verbesserung der Residuen gebracht hat. Für dieses Modell ist auch die verallgemeinerte Methode der kleinsten Quadrate noch etwas besser als die klassische Methode.

Die klassische Methode ist im Allgemeinen ein geeignetes praktisches Verfahren, wenn man von den theoretischen Problemen absieht. Ein klarer Vorteil der verallgemeinerten Methode ist die einfachere Variablenauswahl, einerseits weil die t -Testgrößen korrekt berechnet werden, andererseits aber auch weil die Multikollinearität stark vermindert wird. Diesem Vorteil stehen die numerischen Probleme gegenüber, die bei der Berechnung auftauchen können. Für einen Einsatz in der Praxis müsste dieser Punkt noch wesentlich verbessert werden.

Tabelle 6-26: Zusammenstellung der Residuenplots

		Klassische Methode	Verallgemeinerte Methode
Sin / Cos Temperatur Wärmeleitung	Sin / Cos		
	Temperatur		
	Wärmeleitung		

7 Anwendung auf Staumauer Sambuco

In diesem Kapitel wird das zweite Anwendungsbeispiel behandelt, die Staumauer Sambuco. Wie schon im letzten Kapitel wurde immer nur ein Teil der Daten für die Regression verwendet, der andere Teil dient zur Überprüfung der Prognose. Eine Erklärung zu den Tabellen und Residuenplots ist im vorhergehenden Kapitel, im Abschnitt 6.2 zu finden.

7.1 Beschreibung

7.1.1 Übersicht

Eine Ansicht und ein Querschnitt der Staumauer Sambuco sind in Bild 7-1 gezeigt.

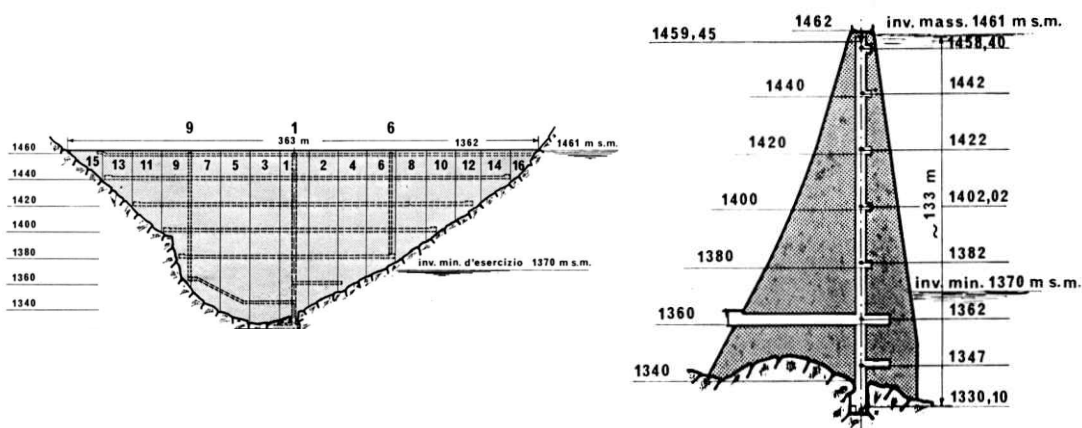


Bild 7-1: Staumauer Sambuco

Messwerte stehen von 1995 bis Mitte 2001 zur Verfügung. Im Gegensatz zu den täglichen Messwerten von Schlegeis gibt es bei Sambuco nur etwa zwei Messtage pro Monat, in nicht ganz regelmässigen Abständen. Da die Mauer 1956 in Betrieb genommen wurde, sind keine Langzeitverformungen zu erwarten. Von den zahlreichen Messungen, die zur Verfügung stehen, wurden die folgenden ausgewählt:

- Pendel MV 1 8 (Verschiebung der Krone im mittleren Bereich der Sperre)
- Stauhöhe: „Quota lago Rittmeyer“
- Alle Auftriebsdrücke (ausser M 6 3 und M 7 3, die praktisch Null sind)
- Alle Betontemperaturen

Die Messstellen für die Betontemperaturen sind in Bild 7-2 gezeigt. Angegeben sind die jeweiligen Ablesepunkte, die Thermometer sind über den Querschnitt verteilt.

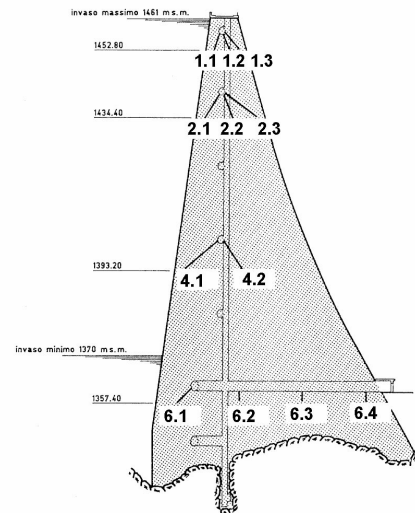


Bild 7-2: Messstellen für Betontemperaturen in Block 1

7.1.2 Beschreibung der Messdaten

Die Zielgrösse ist die Pendelablesung (MV 1 8), die der Kronenverschiebung im mittleren Bereich der Sperre entspricht. Sie ist in Bild 7-3 geplottet. Die Haupteinflussgrösse ist die Stauhöhe, geplottet in Bild 7-4. Weitere Plots der Temperaturen und Auftriebsdrücken finden sich am Ende dieses Kapitels.

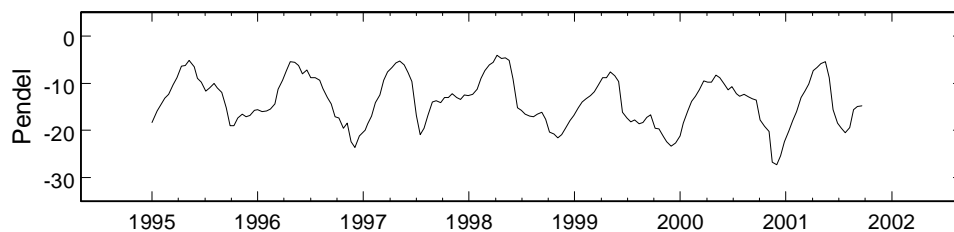


Bild 7-3: Pendel MV 1 8

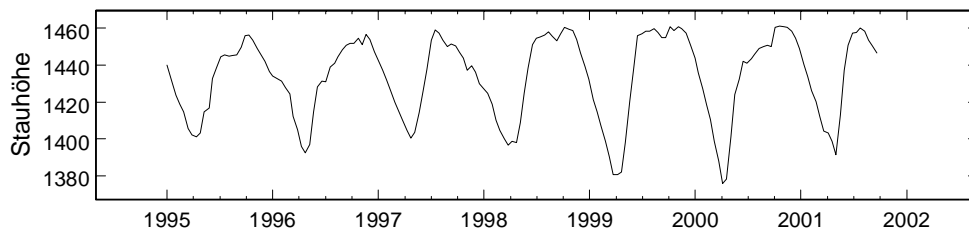


Bild 7-4: Stauhöhe

7.2 Klassisches Modell ohne Temperaturen

Der Ausgangspunkt ist auch hier das klassische Modell mit 4 Tschebyscheff-Polynomen der Stauhöhe, Sinus- und Kosinus-Ansatz für Temperatureffekte und einer Exponentialfunktion für Langzeiteffekte.

7.2.1 Methode der kleinsten Quadrate

Mit der klassischen Methode der kleinsten Quadrate erhält man die Grössen in Tabelle 7-1. Wie daraus ersichtlich ist, können einige Regressoren eliminiert werden. Neben den Regressoren mit p -Werten über 1 % kann man aufgrund des kleinen t -Wertes auch die Exponentialfunktion weglassen. Die schrittweise Elimination führt zum Modell in Tabelle 7-2. Das Modell ist soweit problemlos, ausser dass der Durbin-Watson-Kennwert eine Autokorrelation der Residuen anzeigt. Der berechnete Wert von 0.68 liegt deutlich unter dem unteren Grenzwert von 1.484 (5% Wert für 9 Regressoren und 100 Beobachtungen).

Tabelle 7-1: Ausgangsmodell mit saisonalen Funktionen

Regressor	Coefficient	Std Err	t-stat	p-value	VIF	Std Coeff
Constant	-11.3336	0.2008	-56.43	7.204E-75	0	0
T1(Quota)	-10.0916	0.5335	-18.92	5.572E-34	9.013	-1.15641
T2(Quota)	-2.60644	0.2531	-10.3	3.804E-17	2.167	-0.308726
T3(Quota)	-0.140704	0.2236	-0.6292	0.5307	2.187	-0.0189493
T4(Quota)	-0.120206	0.1876	-0.6408	0.5232	1.706	-0.0170435
sin(s)	-3.00066	0.4464	-6.722	1.327E-9	9.782	-0.428074
cos(s)	-3.94024	0.1527	-25.8	1.078E-44	1.157	-0.565027
sin(2s)	-0.231456	0.1475	-1.569	0.12	1.07	-0.0330451
cos(2s)	-0.403841	0.2351	-1.717	0.08915	2.738	-0.0578666
exp(-t)	-1.26204	0.4148	-3.042	0.003036	1.123	-0.0656441
SS Res	MS Res	R2	F	signif F	DW	Observ
101.081	1.06401	0.96061	257.4	1.241E-62	0.71	105
Regression: 02/01/1995-30/12/1998			Prediction: 14/01/1999-19/09/2001			
SS Res	MS' Res		SS Res	MS' Res		
101.081	0.962677		264.857	3.73038		

Tabelle 7-2: Modell mit saisonalen Funktionen

Regressor	Coefficient	StdErr	t-stat	p-value	VIF	Std Coeff
Constant	-11.3522	0.1437	-79.01	6.212E-92	0	0
T1(Quota)	-10.819	0.4231	-25.57	1.173E-45	5.172	-1.23976
T2(Quota)	-2.23191	0.1994	-11.2	2.42E-19	1.227	-0.264364
sin(s)	-3.6207	0.3489	-10.38	1.473E-17	5.45	-0.516528
cos(s)	-3.94906	0.157	-25.15	4.848E-45	1.115	-0.566292
SS Res	MS Res	R2	F	signif F	DW	Observ
116.64	1.1664	0.95455	525	3.679E-66	0.68	105
Regression: 02/01/1995-30/12/1998			Prediction: 14/01/1999-19/09/2001			
SS Res	MS' Res		SS Res	MS' Res		
116.64	1.11086		317.827	4.47644		

Der Vergleich zwischen Messung und Prognose wird in Bild 7-5 gezeigt. Bild 7-6 zeigt die entsprechenden Residuen. Auffallend sind die Bereiche, wo eine Extrapolation angezeigt wird, vor allem im Frühjahr 1999 und 2000. Dort wurde der Wasserspiegel tiefer abgesenkt als in früheren Jahren, was die relativ schlechte Prognose erklärt. Bild 7-7 zeigt die Verteilung der Residuen. Sowohl die Normalverteilung wie auch die kon-

stante Varianz sind erfüllt. Das Korrelogramm ist in Bild 7-8 gezeigt. Wie schon der Durbin-Watson-Test, zeigt auch das Korrelogramm eine Autokorrelation der Residuen.

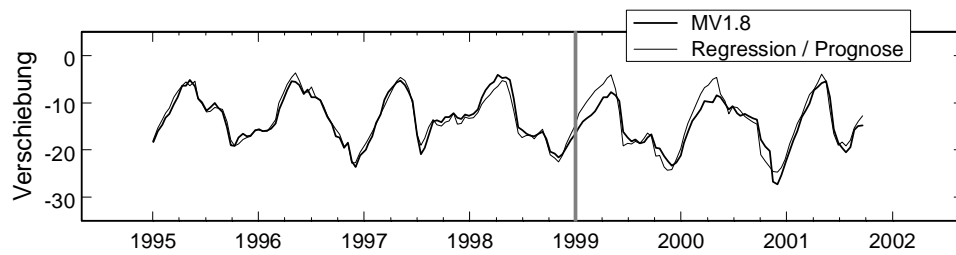


Bild 7-5: Regression, Modell mit saisonalen Funktionen

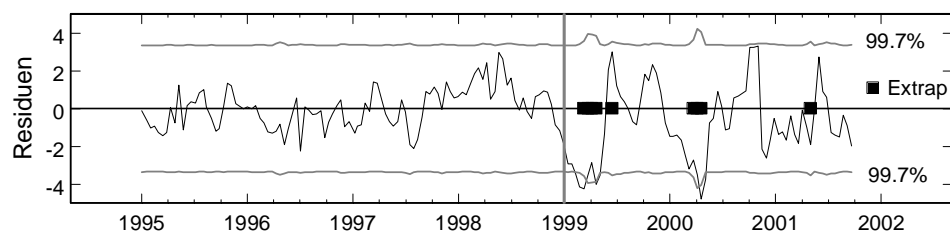


Bild 7-6: Residuen, Modell mit saisonalen Funktionen

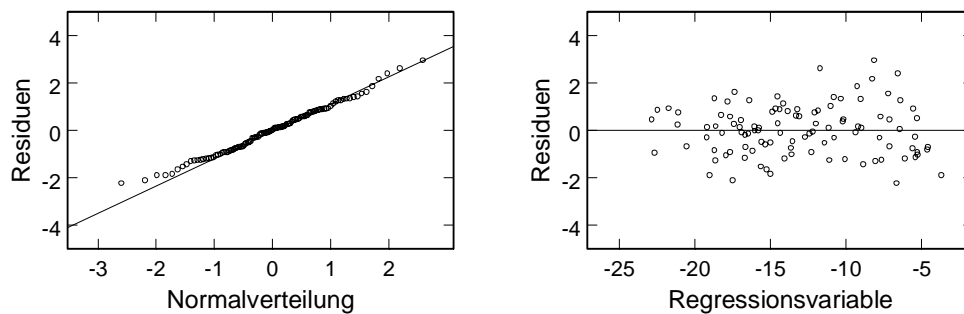


Bild 7-7: Verteilung der Residuen, Modell mit saisonalen Funktionen

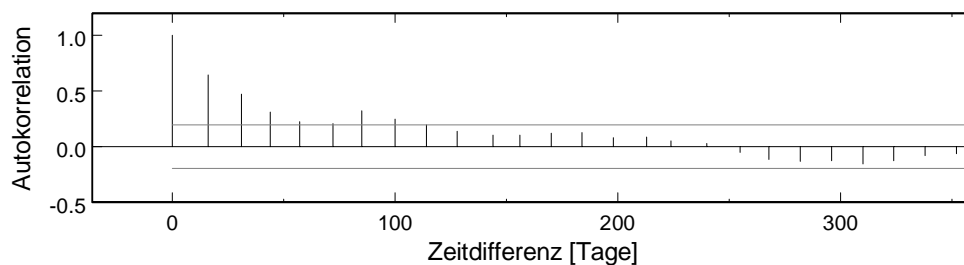


Bild 7-8: Korrelogramm, Modell mit saisonalen Funktionen

7.2.2 Autokorrelation der Fehler

Um die Autokorrelation der Residuen zu berücksichtigen, wird wieder die verallgemeinerte Methode der kleinsten Quadrate verwendet. Der Autokorrelationsparameter wird $\rho = 0.754$. Die Werte für die transformierten Variablen sind in Tabelle 7-3 festgehalten. Die Regressionskoeffizienten und t -Kenngrossen gelten auch für die ursprünglichen Variablen. Wie erwartet sind die t -Kenngrossen gegenüber der gewöhnlichen Methode wesentlich kleiner geworden.

Regressoren mit einem p -Wert über 1% können schrittweise eliminiert werden, was zu einem Modell mit 5 Regressoren führt. Wie Bild 7-9 zeigt, beruht die Prognose im Frühjahr 1999 und 2000 auf einer Extrapolation und die Residuen sind entsprechend gross. Die Details dieses Modells sind hier nicht festgehalten, aber die Quadratsummen sind aus der Zusammenfassung in Tabelle 7-5 ersichtlich.

Tabelle 7-3: Ausgangsmodell mit saisonalen Funktionen, $\rho = 0.754$

Regressor	Coefficient	Std Err	t-stat	p-value	VIF	Std Coeff
Constant	-11.9135	0.4186	-28.46	2.739E-48	0	0
T1(Quota)	-7.9795	0.6573	-12.14	4.951E-21	3.751	-0.783906
T2(Quota)	-2.5568	0.2478	-10.32	3.458E-17	1.344	-0.398745
T3(Quota)	-0.519432	0.1863	-2.789	0.006391	1.277	-0.105069
T4(Quota)	0.072698	0.1453	0.5004	0.6179	1.184	0.0181554
sin(s)	-1.42431	0.606	-2.35	0.02082	3.625	-0.149187
cos(s)	-3.87066	0.3169	-12.22	3.46E-21	1.061	-0.419381
sin(2s)	-0.41006	0.2161	-1.897	0.06081	1.032	-0.0642459
cos(2s)	-1.0452	0.2826	-3.698	0.0003635	1.81	-0.165865
exp(-t)	-0.857042	1.025	-0.8358	0.4054	1.252	-0.0311829
SS Res	MS Res	R2	F	signif F	DW	Observ
52.4924	0.552552	0.89441	89.99	1.628E-42	2.14	105
Regression: 02/01/1995-30/12/1998			Prediction: 14/01/1999-19/09/2001			
SS Res	MS' Res		SS Res	MS' Res		
121.559	1.15771		373.265	5.25725		

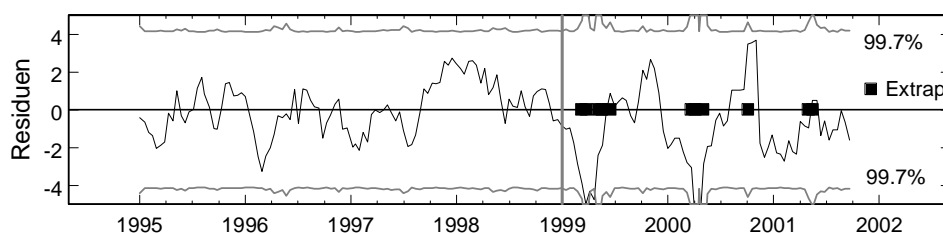


Bild 7-9: Modell mit saisonalen Funktionen, 5 Regressoren, $\rho = 0.815$

Eine wesentlich bessere Prognose erhält man nach der Elimination eines weiteren Regressors. Dieses Modell ist in Tabelle 7-4 festgehalten. Obwohl die Residuen in der Regressionsperiode im Vergleich zur gewöhnlichen Methode grösser werden, wird die Prognose besser. Eine Zusammenfassung wird in Tabelle 7-5 gezeigt. Die Residuenplots sind in der Zusammenfassung am Ende dieses Kapitels nochmals im Vergleich dargestellt.

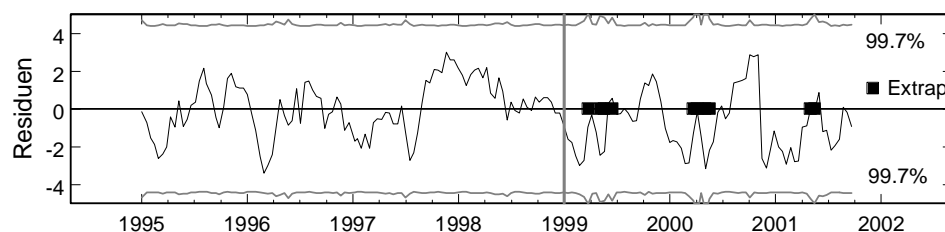
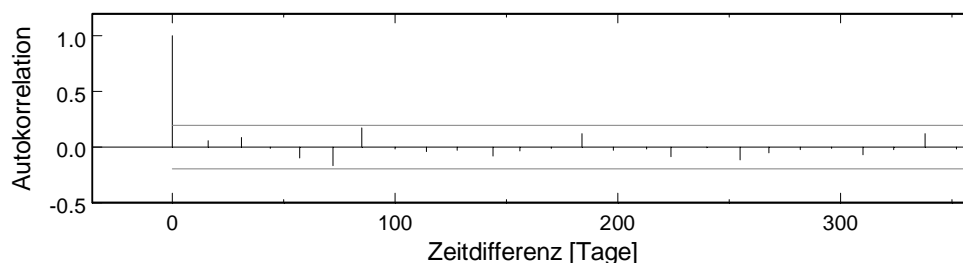
Tabelle 7-4: Modell mit saisonalen Funktionen, 4 Regressoren, $\rho = 0.817$

Regressor	Coefficient	Std Err	t-stat	p-value	VIF	Std Coeff
Constant	-12.3264	0.436	-28.27	1.694E-49	0	0
T1(Quota)	-6.67191	0.4347	-15.35	4.874E-28	1.17	-0.644392
T2(Quota)	-2.48173	0.2628	-9.445	1.634E-15	1.237	-0.407867
cos(s)	-3.76295	0.3883	-9.691	4.718E-16	1.028	-0.381528
cos(2s)	-1.07023	0.2776	-3.855	0.0002045	1.366	-0.17496
SS Res	MS Res	R2	F	signif F	DW	Observ
66.3917	0.663917	0.84929	140.5	3.898E-40	1.88	105
Regression: 02/01/1995-30/12/1998			Prediction: 14/01/1999-19/09/2001			
SS Res	MS' Res	SS Res		MS' Res		
203.907	1.94197	215.466		3.03473		

Tabelle 7-5: Vergleich gewöhnliche und verallgemeinerte Methode (Tabelle 7-2 und Tabelle 7-4 plus Modell mit 5 Regressoren)

		Regression: 02/01/1995-30/12/1998		Prediction: 14/01/1999-19/09/2001	
	Reg	SS Res	MS' Res	SS Res	MS' Res
$\rho = 0$	4	116.64	1.11086	317.827	4.47644
$\rho = 0.817$	4	203.907	1.94197	215.466	3.03473
$\rho = 0.815$	5	176.888	1.68465	361.42	5.09043

Die graphische Darstellung der Residuen ist in Bild 7-10 gezeigt. Auch hier fallen die Extrapolationsbereiche in der Prognoseperiode auf, aber sie sind weniger ausgeprägt als im Modell mit 5 Regressoren (Bild 7-9). Bild 7-11 zeigt das Korrelogramm der Residuen der transformierten Variablen.

**Bild 7-10: Residuen, Modell mit saisonalen Funktionen, 4 Regressoren, $\rho = 0.817$** **Bild 7-11: Korrelogramm, Modell mit saisonalen Funktionen, 4 Regressoren, $\rho = 0.817$**

7.2.3 Abhängigkeit von der Regressionsperiode

In den bisher betrachteten Modellen fällt auf, dass die Prognose wesentlich schlechter ist als die Regression. Da stellt sich die Frage, ob das an der Prognose selber liegt, oder ob sich einfach die Mauer in einer Periode ganz anders verhält als in der andern. Wie sich zeigt, gibt hier eine Regression über die ganze Messzeit ein gutes Bild über das Verhalten der Mauer in den einzelnen Zeitabschnitten. Die klassische Methode der kleinsten Quadrate liefert ein Modell mit zwei Polynomen und trigonometrischen Funktionen mit einer Periode von einem ganzen und einem halben Jahr. Die Residuen sind in Bild 7-12 geplottet. Man sieht deutlich, dass die Residuen ab etwa 1998 wesentlich grösser werden als vorher. Man muss also auch für eine Prognose ab diesem Zeitpunkt eine wesentlich schlechtere Qualität erwarten.

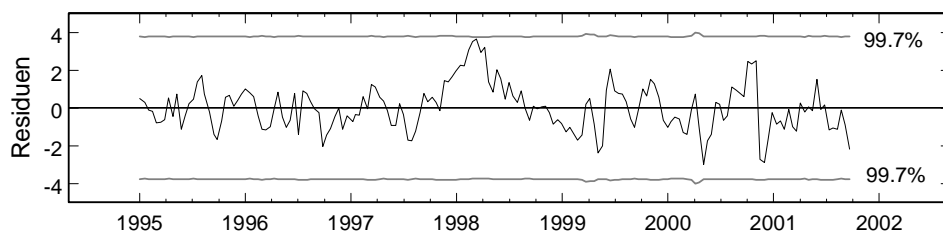


Bild 7-12: Residuen, Modell mit saisonalen Funktionen, lange Regressionsperiode

7.3 Modell mit Betontemperaturen

7.3.1 Methode der kleinsten Quadrate

Als weitere Regressoren werden die gemessenen Betontemperaturen ins Modell einbezogen. Die Sinus- und Kosinusterme, die den Einfluss der Temperatur modellieren sollen, werden somit weggelassen. Dasselbe gilt für die Exponentialfunktion. Das volle Modell hat damit 16 Regressoren. Der maximale VIF-Wert ist mit 80 nicht allzu hoch, sodass eine schrittweise Elimination erfolgreich sein dürfte. Es bleiben 5 Regressoren und ein maximaler VIF-Wert von 16, der durch eine Principal-Component-Regression beseitigt wird. Das Modell wird in Tabelle 7-6 gezeigt.

Tabelle 7-6: Modell mit Betontemperaturen, ein Eigenvektor weggelassen

Regressor	Coefficient	Std Err	t-stat	p-value	VIF	Std Coeff
Constant	-2.84761	1.491	-1.91	0.05905	0	0
T1(Quota)	-9.11379	0.2544	-35.82	1.766E-58	2.337	-1.04436
T2(Quota)	-2.48009	0.1817	-13.65	1.724E-24	1.273	-0.29376
T 1 2	-0.186155	0.02401	-7.752	8.129E-12	0.4059	-0.0941801
T 2 2	-2.15336	0.1227	-17.55	3.795E-32	0.7383	-0.287508
T 2 3	1.13135	0.0472	23.97	5.218E-43	1.573	0.57332
SS Res	MS Res	R2	F	signif F	DW	Observ
92.3947	0.93328	0.96400	530.1	8.93E-70	0.94	105
Regression: 02/01/1995-30/12/1998			Prediction: 14/01/1999-19/09/2001			
SS Res	MS' Res		SS Res	MS' Res		
92.3947	0.87995		222.106	3.12826		

Die Autokorrelation der Residuen ist nicht sehr ausgeprägt, aber der Durbin-Watson liegt doch unter dem unteren Grenzwert von 1.571 (5% Wert für 5 Regressoren und 100 Beobachtungen).

Der Verlauf der Residuen ist in Bild 7-13 geplottet. Wieder fallen die Bereiche mit einer Extrapolation auf. Bild 7-14 zeigt das Korrelogramm.

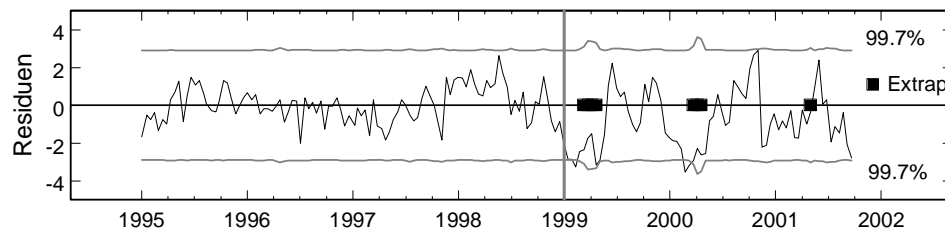


Bild 7-13: Residuen, Modell mit Betontemperaturen, ein Eigenvektor weggelassen

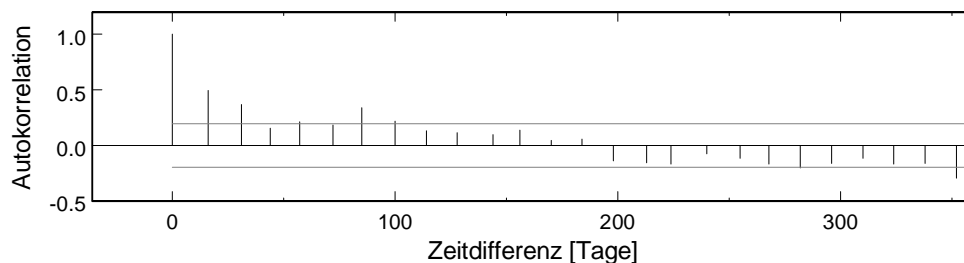


Bild 7-14: Korrelogramm, Modell mit Betontemperaturen, ein Eigenvektor weggelassen

7.3.2 Autokorrelation der Fehler

Die verallgemeinerte Methode der kleinsten Quadrate führt für das volle Modell auf einen maximalen VIF-Wert von etwa 14. Werden alle nichtsignifikanten Regressoren eliminiert, bleiben 4 Regressoren, wie in Tabelle 7-7 gezeigt. Der Autokorrelationsparameter wird $\rho = 0.851$.

Tabelle 7-7: Modell mit Betontemperaturen, $\rho = 0.851$

Regressor	Coefficient	Std Err	t-stat	p-value	VIF	Std Coeff
Constant	-5.84221	3.225	-1.812	0.07305	0	0
T1(Quota)	-8.11824	0.4118	-19.72	3.133E-36	1.08	-0.780371
T2(Quota)	-2.23304	0.2295	-9.729	3.897E-16	1.039	-0.377635
T 1 3	0.572971	0.067	8.552	1.455E-13	1.262	0.365953
T 2 2	-1.36793	0.2959	-4.622	1.136E-5	3.454	-0.327171
SS Res	MS Res	R2	F	signif F	DW	Observ
60.4041	0.604041	0.85495	142.1	2.456E-40	2.19	105
Regression: 1995-01-02-1998-12-30			Prediction: 14/01/1999-19/09/2001			
SS Res	MS' Res	SS Res		MS' Res		
193.256	1.84053	246.624		3.47357		

Sowohl in der Regressionsperiode wie auch in der Prognoseperiode sind die Residuen grösser als bei der gewöhnlichen Methode, wie die Zusammenfassung in Tabelle 7-8 zeigt.

Die Verschlechterung in der Prognose ist vor allem auf die Verschiebung der Nulllinie zurückzuführen, wie der Plot der Residuen (Bild 7-15) zeigt. Die Amplitude selber ist eher kleiner. Der Langzeittrend der Residuen ist auch schon in Bild 7-13 zu sehen, wenn auch weniger ausgeprägt. Das Korrelogramm für die Residuen des transformierten Problems ist in Bild 7-16 gezeigt.

Tabelle 7-8: Vergleich gewöhnliche und verallgemeinerte Methode (Tabelle 7-6 und Tabelle 7-7)

		Regression: 02/01/1995-30/12/1998		Prediction: 14/01/1999-19/09/2001	
	Reg	SS Res	MS' Res	SS Res	MS' Res
$\rho = 0$	5	92.3947	0.87995	222.106	3.12826
$\rho = 0.851$	4	193.256	1.84053	246.624	3.47357

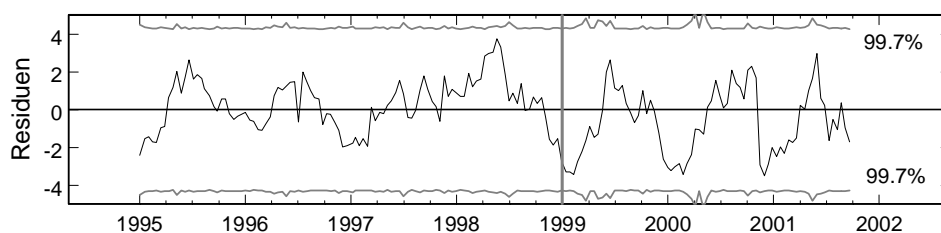


Bild 7-15: Residuen, Modell mit Betontemperaturen, $\rho = 0.851$

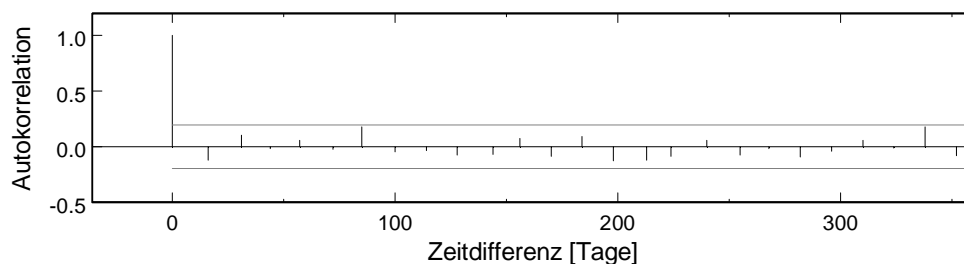


Bild 7-16: Korrelogramm, Modell mit Betontemperaturen, $\rho = 0.851$

7.4 Modell mit Auftriebsdrücken

7.4.1 Methode der kleinsten Quadrate

Neben den Betontemperaturen stehen auch die Auftriebsdrücke zur Verfügung. Beginnt man mit einem Modell, das alle Betontemperaturen und alle Auftriebsdrücke enthält, so stellt man wieder viele VIF-Werte über 10, zum Teil bis fast 1000, fest. Eliminiert man schrittweise alle Regressoren mit einem p -Wert über 1%, so erhält man ein Modell mit 6 Regressoren, bei dem man noch einen Eigenvektor entfernen muss, um die VIF-Werte unter 10 zu halten. Die Elimination mit einem Verzerrungsparameter von $k = 0.005$

führt zum gleichen Modell. Das Resultat ist in Tabelle 7-9 festgehalten. Der Durbin-Watson-Kennwert von 1.47 ist gerade knapp unter dem unteren Grenzwert von 1.518 (5% Wert für 5 Regressoren und 90 Beobachtungen). Theoretisch müsste er eigentlich über dem oberen Grenzwert von 1.776 liegen, damit eine Autokorrelation ausgeschlossen werden könnte. Das Korrelogramm in Bild 7-18 zeigt ebenfalls nur noch eine schwache Autokorrelation, die toleriert werden kann. Die Residuen sind in Bild 7-17 geplottet.

Interessant ist, dass der Auftrieb zum Teil die Polynome ersetzt hat und auch die Autokorrelation der Residuen beseitigt hat. Der Auftrieb M61 verläuft praktisch gleich wie die Stauhöhe.

Tabelle 7-9: Modell mit Betontemperaturen und Auftrieb, ein Eigenvektor weggelassen

Regressor	Coefficient	Std Err	t-stat	p-value	VIF	Std Coeff
Constant	-3.67961	1.601	-2.298	0.02371	0	0
T2(Quota)	-0.905582	0.1734	-5.223	9.891E-7	1.77	-0.107264
M 2 3	-0.963851	0.03214	-29.99	3.549E-51	1.22	-0.511257
M 6 1	-0.136717	0.004243	-32.22	6.031E-54	1.109	-0.523692
T 2 1	-0.457749	0.09315	-4.914	3.582E-6	6.436	-0.192409
T 2 2	-1.14993	0.1425	-8.069	1.808E-12	1.52	-0.153535
T 2 3	1.12397	0.06085	18.47	1.095E-33	3.992	0.569583
SS Res	MS Res	R2	F	signif F	DW	Observ
59.9098	0.611325	0.97665	683.3	1.342E-77	1.47	105
Regression: 02/01/1995-30/12/1998			Prediction: 14/01/1999-19/09/2001			
SS Res	MS' Res	SS Res		MS' Res		
59.9098	0.57057	142.261		2.00367		

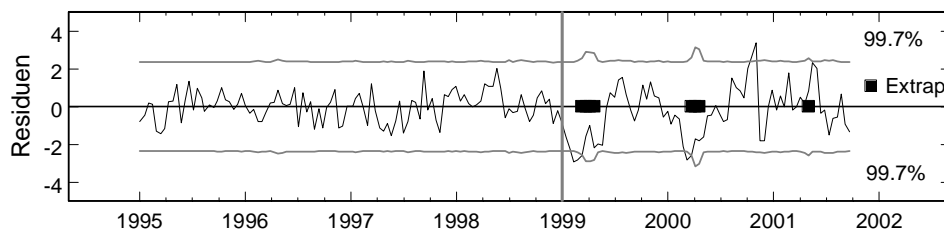


Bild 7-17: Residuen, Auftrieb und Betontemperaturen

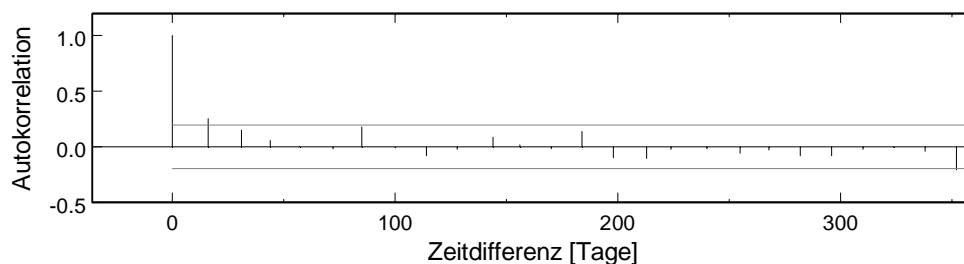


Bild 7-18: Korrelogramm, Auftrieb und Betontemperaturen

7.4.2 Autokorrelation der Fehler

Die Autokorrelation der Fehler war im letzten Modell zwar relativ gering, aber es ist dennoch interessant zu sehen, wie in diesem Fall die Lösung mit der verallgemeinerten Methode der kleinsten Quadrate aussieht. Für das ganze Modell ist der maximale VIF-Wert etwa 130. Nach der Elimination der Regressoren mit p -Werten über 1% bleiben 6 Regressoren im Modell. Mit der Elimination von 2 weiteren Regressoren sind auch alle VIF-Werte unter 10. Das Resultat wird noch etwas verbessert, wenn man nochmals zwei Regressoren einfügt und wieder zwei eliminiert. Das Resultat ist in Tabelle 7-10 festgehalten. Die Residuen sind etwas kleiner als in den Modellen ohne Auftrieb, aber wesentlich grösser als bei der klassischen Lösung für das Modell mit Auftrieb (Tabelle 7-9). Wie der Plot der Residuen in (Bild 7-19) zeigt, ist die grössere Abweichung in der Prognoseperiode wieder auf eine Verschiebung der Nulllinie zurückzuführen. Das Korrelogramm für die Residuen des transformierten Problems ist in (Bild 7-20) gezeigt.

Tabelle 7-10: Modell mit Betontemperaturen und Auftrieb, $\rho = 0.722$

Regressor	Coefficient	Std Err	t-stat	p-value	VIF	Std Coeff
Constant	4.62896	2.873	1.611	0.1102	0	0
T2(Quota)	-1.05599	0.2272	-4.648	1.024E-5	1.113	-0.160525
M 6 1	-0.252582	0.01079	-23.41	2.303E-42	1.307	-0.875951
T 1 3	0.545635	0.06073	8.985	1.661E-14	1.395	0.347361
T 2 2	-1.52754	0.2658	-5.748	9.837E-8	3.679	-0.360854
SS Res	MS Res	R2	F	signif F	DW	Observ
57.0042	0.570042	0.89287	202.6	4.978E-47	2.32	105
Regression: 02/01/1995-30/12/1998			Prediction: 14/01/1999-19/09/2001			
SS Res	MS' Res	SS Res		MS' Res		
112.36	1.07009	199.654		2.81203		

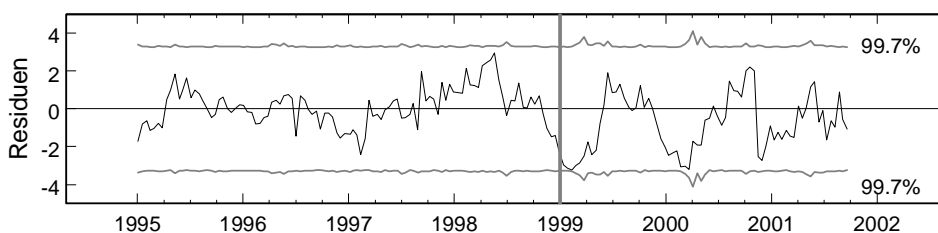


Bild 7-19: Residuen, Modell mit Betontemperaturen und Auftrieb, $\rho = 0.722$

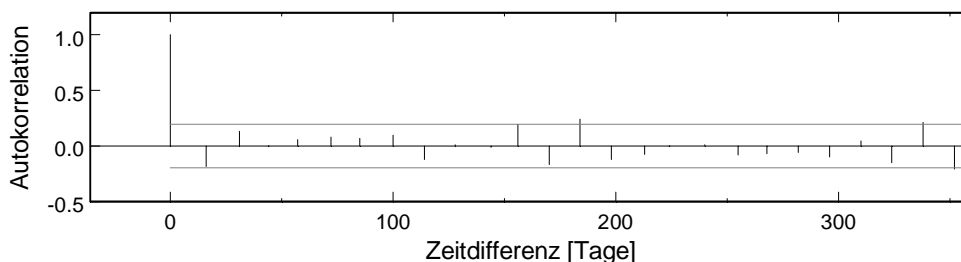


Bild 7-20: Korrelogramm, Modell mit Betontemperaturen und Auftrieb, $\rho = 0.722$

7.5 Modell mit Wärmeleitung

Will man auch noch die Wärmeleitung einbeziehen, kann man dies mit oder ohne Auftriebsdrücke tun. Um nicht allzu viele Parameter gleichzeitig zu variieren, werden hier nur die gemessenen und die berechneten Temperaturen benützt und der Auftrieb weggelassen. Bei der speziellen Querschnittsform der Mauer ist die eindimensionale Theorie der Wärmeleitung höchstens im oberen Teil anwendbar, wo die Mauer nicht zu dick ist. Bei einem Zeitschritt von etwa 15 Tagen ist die Auflösung der Messungen für eine Berechnung der Temperatur in 2 m Tiefe zu klein (siehe Kapitel 5). Aus diesen Überlegungen heraus werden die Temperaturen auf dem obersten Niveau in 4 m Tiefe und auf dem zweitobersten Niveau in 4 und 6 m Tiefe berechnet.

7.5.1 Methode der kleinsten Quadrate

Das Ausgangsmodell hat 12 gemessene und 6 berechnete Temperaturen. Obwohl der höchste VIF-Wert bei 1000 liegt, funktioniert die schrittweise Elimination gut. Mit 8 Regressoren sind alle p-Werte unter 1%, aber der höchste VIF ist immer noch 37. Man muss zwei weitere Regressoren eliminieren, um die VIF-Werte unter 10 zu bekommen. Das Resultat ist in Tabelle 7-11 zu sehen. Wenn man noch einen weiteren Regressor eliminiert, erhält man eine wesentlich bessere Prognose, das auf Kosten von grösseren Residuen in der Regressionsperiode (Tabelle 7-12).

Tabelle 7-11: Modell mit Wärmeleitung, 5 Regressoren

Regressor	Coefficient	Std Err	t-stat	p-value	VIF	Std Coeff
Constant	-20.8049	0.4259	-48.85	1.665E-64	0	0
T1(Quota)	-8.54881	0.2182	-39.18	1.279E-56	2.398	-0.964552
T2(Quota)	-2.46646	0.1486	-16.6	1.482E-28	1.249	-0.294989
T 1 3	0.717796	0.0352	20.39	1.076E-34	2.251	0.486417
temp(T 1 1 4)	-3.39267	0.247	-13.73	2.146E-23	8.209	-0.625524
temp(T 2 3 4)	2.60928	0.309	8.444	6.653E-13	8.545	0.392406
SS Res	MS Res	R2	F	signif F	DW	Observ
49.3675	0.57404	0.97826	774.1	6.751E-70	1.05	92
Regression: 05/07/1995-30/12/1998			Prediction: 14/01/1999-19/09/2001			
SS Res	MS' Res		SS Res	MS' Res		
49.3675	0.536603		204.883	2.88568		

Tabelle 7-12: Modell mit Wärmeleitung, 4 Regressoren

Regressor	Coefficient	Std Err	t-stat	p-value	VIF	Std Coeff
Constant	-21.7817	0.5512	-39.52	2.315E-57	0	0
T1(Quota)	-8.3519	0.2917	-28.63	4.696E-46	2.371	-0.942335
T2(Quota)	-2.30852	0.1982	-11.65	1.908E-19	1.23	-0.2761
T 1 3	0.804953	0.04525	17.79	9.987E-31	2.058	0.54548
temp(T 1 1 4)	-1.50574	0.1416	-10.63	2.101E-17	1.492	-0.27762
SS Res	MS Res	R2	F	signif F	DW	Observ
90.2992	1.03792	0.96024	525.3	5.075E-60	0.66	92
Regression: 05/07/1995-30/12/1998			Prediction: 14/01/1999-19/09/2001			
SS Res	MS' Res		SS Res	MS' Res		
90.2992	0.981513		157.751	2.22184		

Die Residuenplots für beide Fälle sind in Bild 7-21 und Bild 7-22 gezeigt. Der Unterschied in der Prognose liegt vor allem wieder im Langzeitverhalten.

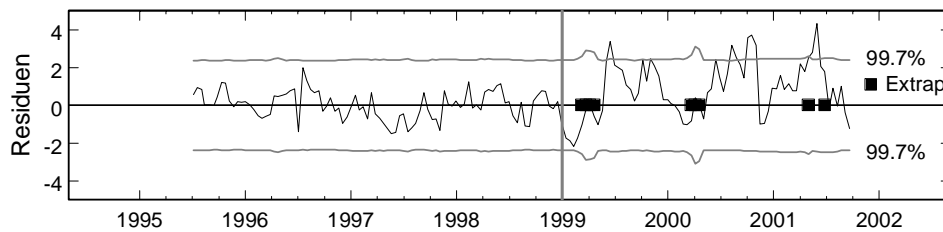


Bild 7-21: Modell mit Wärmeleitung, 5 Regressoren

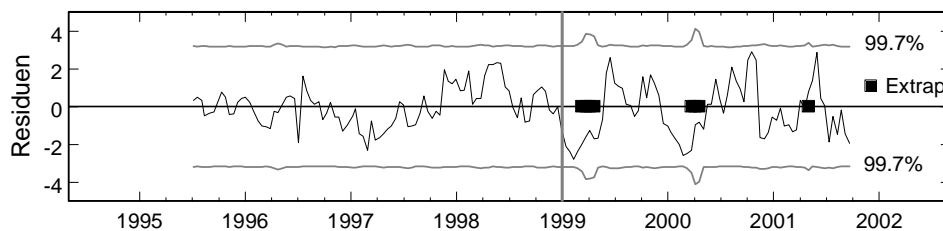


Bild 7-22: Modell mit Wärmeleitung, 4 Regressoren

7.5.2 Autokorrelation der Fehler

Der Durbin-Watson-Kennwert von 1 oder weniger deutet auch hier auch Autokorrelation der Fehler hin. Darum wird hier auch noch eine Berechnung mit der verallgemeinerten Methode der kleinsten Quadrate gezeigt. Die schrittweise Elimination der nicht-signifikanten Regressoren führt zu 6 Regressoren, ein zusätzlicher muss eliminiert werden, um all VIF-Werte unter 10 zu bringen. Der Autokorrelationsparameter im Endmodell ist $\rho = 0.552$. Das Resultat ist in Tabelle 7-13 gezeigt. Wie bei der klassischen Methode wird auch hier die Prognose wesentlich besser, wenn ein weiterer Regressor eliminiert wird. Diese Resultate sind in Tabelle 7-14 gezeigt.

Tabelle 7-13: Modell mit Wärmeleitung, 5 Regressoren, $\rho = 0.552$

Regressor	Coefficient	Std Err	t-stat	p-value	VIF	Std Coeff
Constant	-19.7965	0.6132	-32.28	7.605E-50	0	0
T1(Quota)	-8.13478	0.3242	-25.09	2.488E-41	1.834	-0.876565
T2(Quota)	-2.29334	0.1915	-11.97	5.179E-20	1.124	-0.32755
T 1 3	0.630854	0.04961	12.72	1.848E-21	1.848	0.446049
temp(T 1 1 4)	-3.6533	0.4182	-8.735	1.702E-13	7.569	-0.620032
temp(T 2 3 4)	2.79317	0.5337	5.234	1.159E-6	8.106	0.384455
SS Res	MS Res	R2	F	signif F	DW	Observ
37.0814	0.431179	0.94276	285.3	5.851E-52	2.11	92
Regression: 05/07/1995-30/12/1998			Prediction: 14/01/1999-19/09/2001			
SS Res	MS' Res		SS Res	MS' Res		
53.4035	0.580473		195.967	2.7601		

Tabelle 7-14: Modell mit Wärmeleitung, 4 Regressoren, $\rho = 0.762$

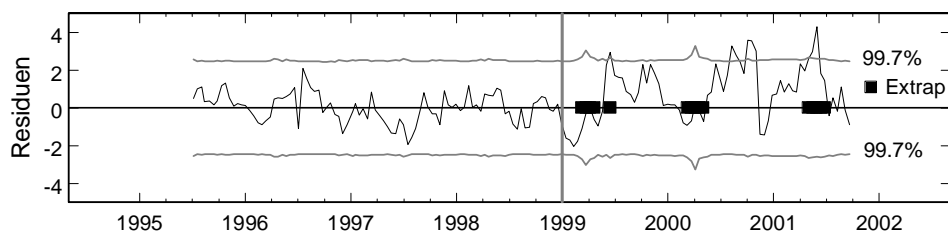
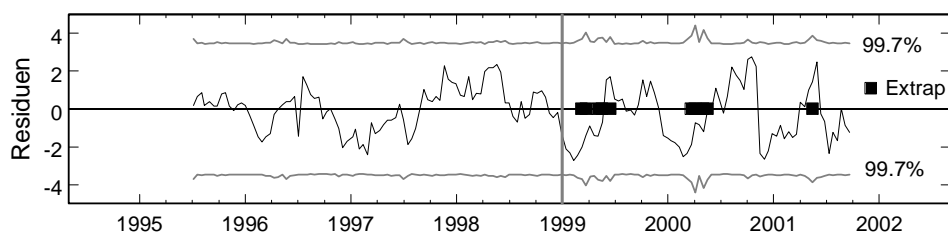
Regressor	Coefficient	Std Err	t-stat	p-value	VIF	Std Coeff
Constant	-19.692	0.8072	-24.4	1.133E-40	0	0
T1(Quota)	-7.45575	0.4187	-17.81	9.364E-31	1.343	-0.749785
T2(Quota)	-2.13328	0.222	-9.609	2.545E-15	1.044	-0.35663
T 1 3	0.620046	0.06116	10.14	2.109E-16	1.456	0.444514
temp(T 1 1 4)	-1.74036	0.2725	-6.386	8.057E-9	1.183	-0.252341
SS Res	MS Res	R2	F	signif F	DW	Observ
45.0816	0.518179	0.88516	168.3	4.428E-40	2.16	92
Regression: 05/07/1995-30/12/1998				Prediction: 14/01/1999-19/09/2001		
SS Res	MS' Res		SS Res	MS' Res		
108.224	1.17634		163.72	2.30591		

Die Resultate der Modelle mit 4 und 5 Regressoren, berechnet mit der klassischen und der verallgemeinerten Methode, sind in Tabelle 7-15 zusammengestellt.

Tabelle 7-15: Vergleich klassische und verallgemeinerte Methode (Tabelle 7-11, Tabelle 7-12, Tabelle 7-13, Tabelle 7-14)

		Regression: 05/07/1995-30/12/1998		Prediction: 14/01/1999-19/09/2001	
	Reg	SS Res	MS' Res	SS Res	MS' Res
$\rho = 0$	5	49.3675	0.536603	204.883	2.88568
$\rho = 0$	4	90.2992	0.981513	157.751	2.22184
$\rho = 0.552$	5	53.4035	0.580473	195.967	2.7601
$\rho = 0.762$	4	108.224	1.17634	163.72	2.30591

Die Residuen sind in Bild 7-23 und Bild 7-24 geplottet. Die Resultate sind sehr ähnlich wie jene der klassischen Methode.

**Bild 7-23: Residuen, Modell mit Wärmeleitung, 5 Regressoren, $\rho = 0.552$** **Bild 7-24: Residuen, Modell mit Wärmeleitung, 4 Regressoren, $\rho = 0.762$**

7.6 Zusammenfassung Sambuco

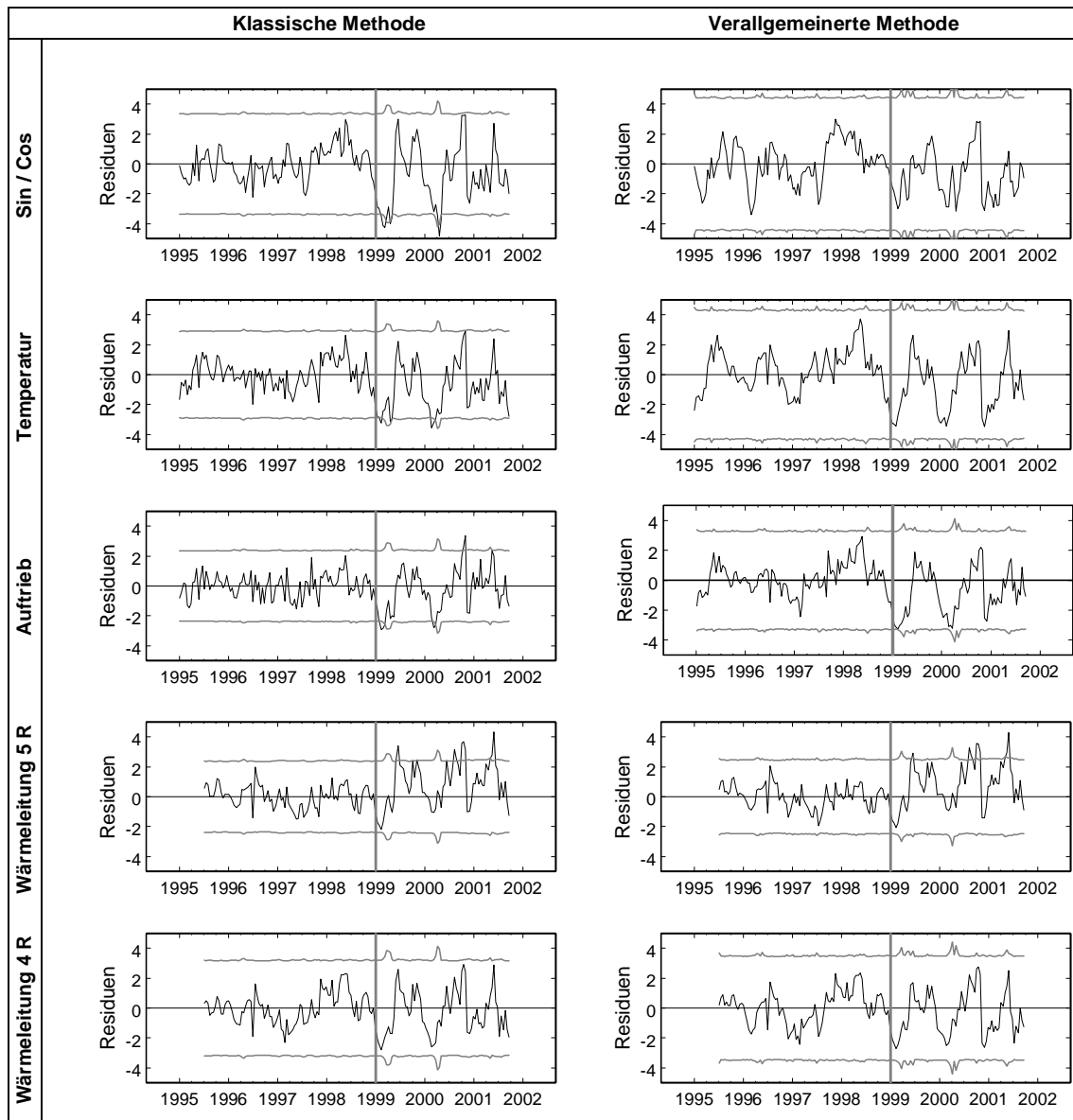
Bei Modellen mit vielen Regressoren wurde die Variablenauswahl durch die Multikollinearität erschwert. In einigen Fällen führte die schrittweise Elimination zum Ziel, in andern Fällen wurde die Auswahl durch eine Ridge-Regression verbessert. Verschiedene Varianten sollten in diesen Fällen ausprobiert werden. Meistens war nach der Reduktion des Modells auch die Multikollinearität verschwunden. Dort, wo das nicht der Fall war, konnte sie durch die Elimination weiterer Regressoren oder mit Hilfe der Principal-Component-Regression beseitigt werden.

Die Autokorrelation der Fehler war bei diesen Messwerten kleiner als beim Beispiel Schlegeis im letzten Kapitel, wahrscheinlich wegen den grösseren Zeitschritten in der Messung. Trotzdem war sie vorhanden und konnte mit der verallgemeinerten Methode der kleinsten Quadrate behandelt werden. Vom Praktischen her gesehen, hat das wenig Vorteile gebracht. Einzig die t -Kenngrössen waren aussagekräftiger und konnten direkt für die Variablenauswahl benützt werden. Die theoretisch kleinere Varianz der Regressionskoeffizienten und die damit verbesserte Prognose konnte aber nicht beobachtet werden.

Als Überblick sind die Residuenplots in Tabelle 7-16 nochmals zusammengestellt. Die graphische Darstellung ist für den Vergleich besser geeignet als die Quadratsummen der Residuen. Insbesondere kann man so noch zwischen der Grösse der Amplitude und einem Abdriften des Mittelwertes unterscheiden. Grundsätzlich sind die Unterschiede zwischen den Modellen und den Methoden nicht so gross wie erwartet. Die grossen Unterschiede zwischen Regression und Prognose sind wahrscheinlich eher auf das unterschiedliche Verhalten der Mauer in beiden Zeitperioden zurückzuführen als auf die Regression selber.

Die aufwändigeren Modelle (von oben nach unten in der Tabelle) liefern tendenziell eine bessere Prognose, aber die Unterschiede sind nicht markant. Die beiden Methoden (links und rechts in der Tabelle) liefern bei allen Modellen—ausser beim Sinus-Kosinus-Modell—vergleichbare Fehleramplituden. Die Unterschiede liegen vor allem im Langzeitverhalten. Die Langzeitabweichungen scheinen sich aber eher zufällig auf die verschiedenen Modelle und Methoden zu verteilen und hängen oft nur davon ab, ob ein Regressor mehr oder weniger im Modell gelassen wird. Dies ist besonders deutlich bei der Wärmeleitung wo zuunterst noch das Modell mit 4 statt 5 Regressoren angefügt ist. Tendenziell scheint es auf einen Kompromiss zwischen der Langzeitanpassung in der Regressionsperiode und der Prognoseperiode hinauszulaufen.

Tabelle 7-16: Zusammenstellung der Residuenplots



7.7 Anhang: Plots der Temperaturen und Auftriebsdrücke

Die Betontemperaturen in den verschiedenen Höhen sind in Bild 7-25, Bild 7-26, Bild 7-27 und Bild 7-28 geplottet. Die Betontemperaturen zuoberst auf Niveau 1 sind alle drei ähnlich, da die Mauer dort relativ dünn ist. Auf dem Niveau 2 sieht man deutlich den Effekt der Wärmeleitung von aussen nach innen. Die Temperaturschwankungen auf Niveau 4 und 6 sind ziemlich klein, ausser die Temperatur T6.4 nahe der Oberfläche an der Luftseite.

Die Auftriebsdrücke sind in Bild 7-29, Bild 7-30, Bild 7-31 und Bild 7-32 geplottet. Die Position der Messpunkte war für diesen Bericht nicht bekannt. Die Drücke M 6 3 und M 7 3 wurden nicht berücksichtigt, da sie praktisch Null sind.

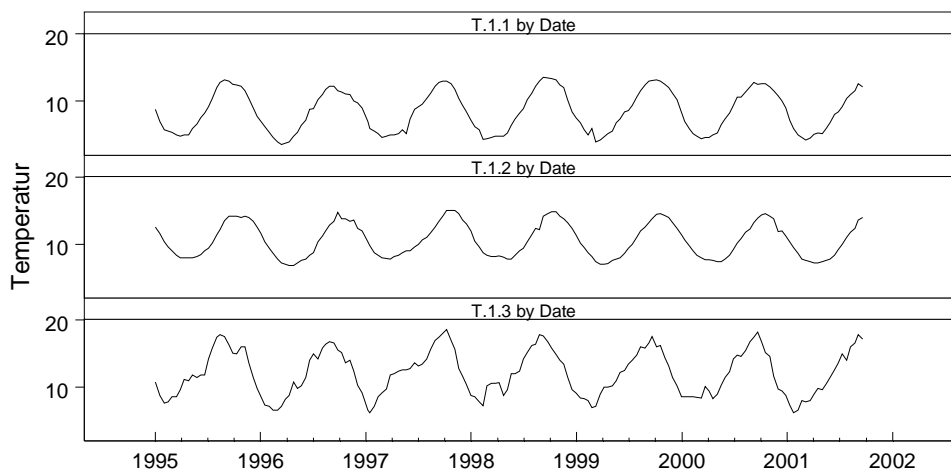


Bild 7-25: Betontemperaturen T 1 1, T 1 2, T 1 3

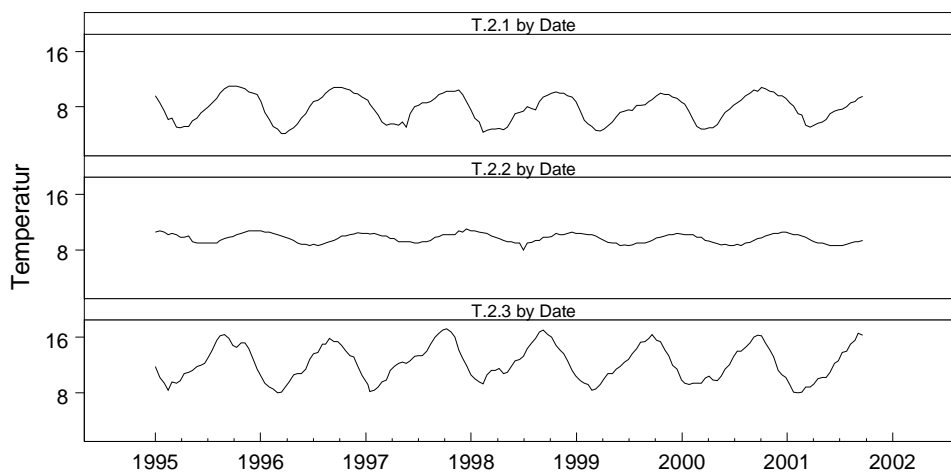


Bild 7-26: Betontemperaturen T 2 1, T 3 2, T 2 3

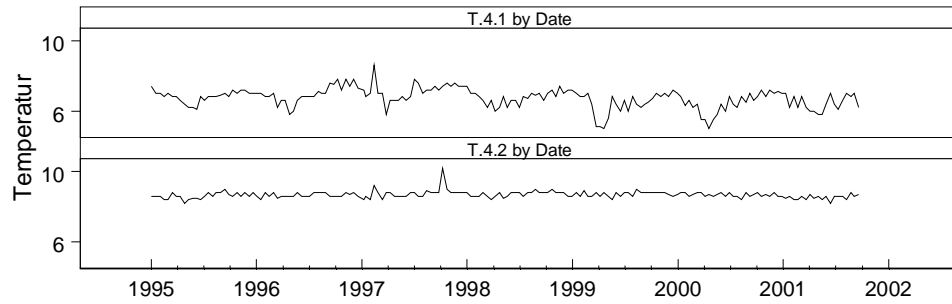


Bild 7-27: Betontemperaturen T 4 1, T 4 2

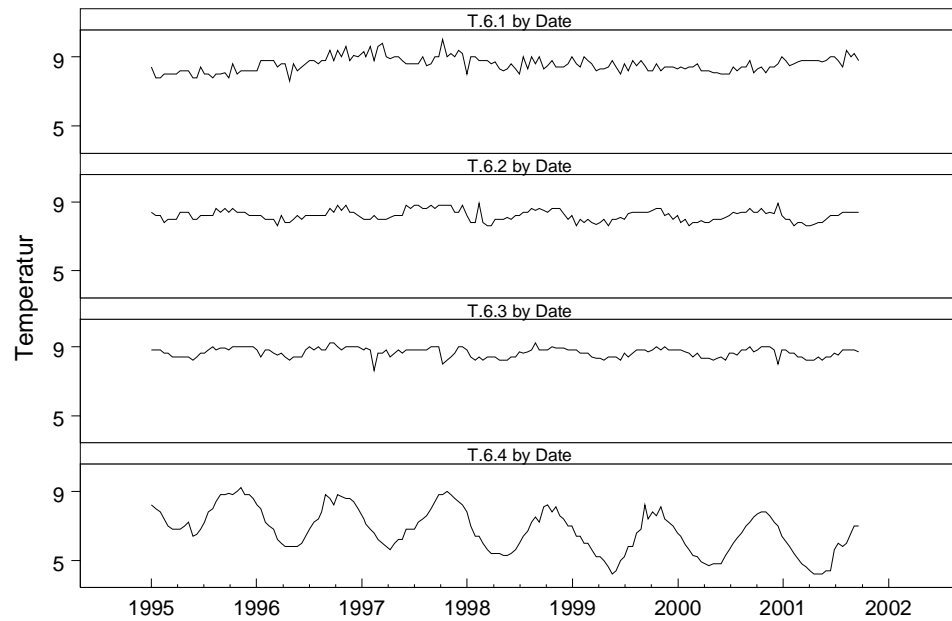


Bild 7-28: Betontemperaturen T 6 1, T 6 2, T 6 3, T 6 4

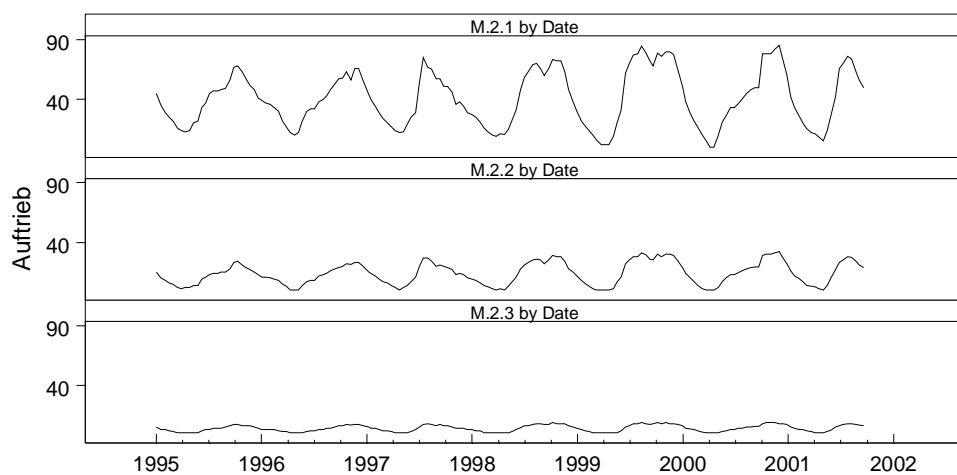


Bild 7-29: Auftrieb M 2 1, M 2 2, M 2 3

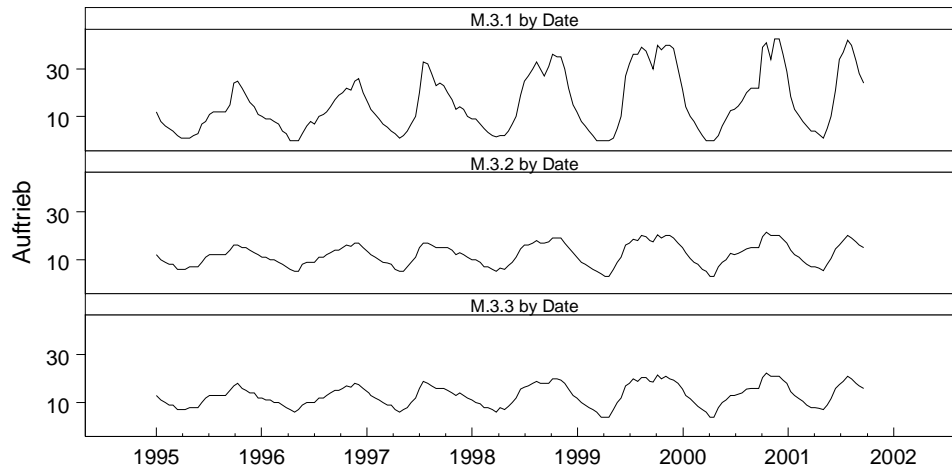


Bild 7-30: Auftrieb M 3 1, M 3 2, M 3 3

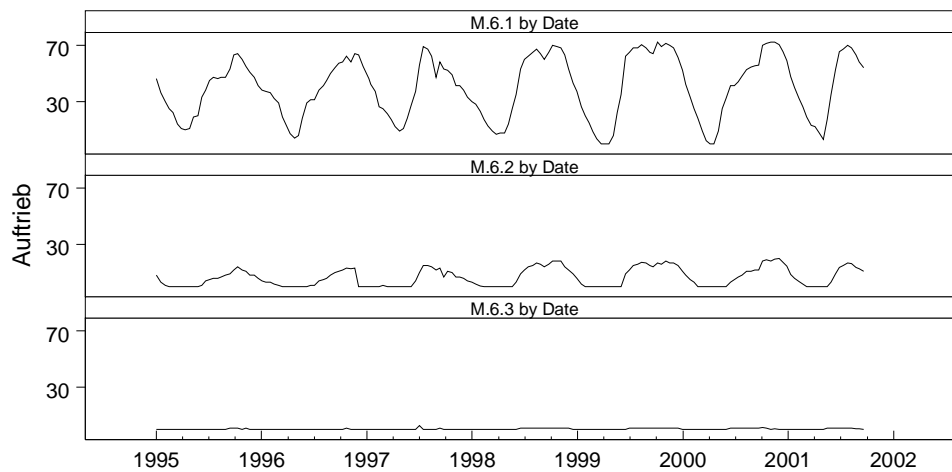


Bild 7-31: Auftrieb M 6 1, M 6 2, M 6 3

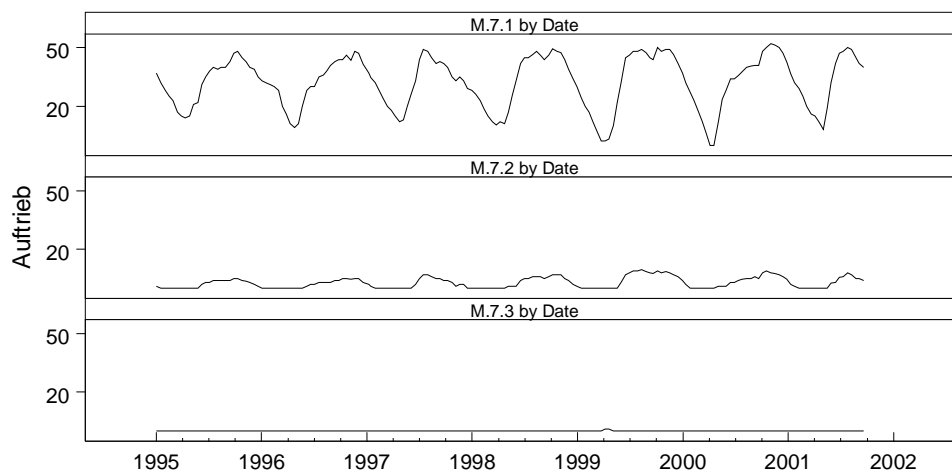


Bild 7-32: Auftrieb M 7 1, M 7 2, M 7 3

8 Beispiel mit Langzeitverformungen

8.1 Beschreibung

Dieses Beispiel wurde bereits in der Semesterarbeit Antille (2001) verwendet. Die Angaben über die Mauer und die Messpunkte fehlen, sodass nur statistische Überlegungen gemacht werden können. Zudem sind keine Temperaturmesswerte vorhanden, sodass das Beispiel weniger interessant ist als die früheren. Eine neue Komponente ergibt sich aber bei den Langzeiteffekten.

Hier wird der radiale Ausschlag des Pendels A betrachtet. Mit den vorhandenen Messdaten kann nur ein Modell mit Stauhöhe und saisonalen Funktionen untersucht werden. Ausserdem spielen auch die Langzeiteffekte eine wichtige Rolle, so dass auch die Exponentialterme berücksichtigt werden müssen. Der Pendelausschlag ist in Bild 8-1 und die Stauhöhe in Bild 8-2 geplottet. Es ist schon aus diesen Bildern ersichtlich, dass weder eine Funktion der Stauhöhe noch eine periodische Funktion zu einer guten Regression führen wird. Während die Stauhöhe jedes Jahr auf ungefähr den gleichen Maximalwert steigt, ist der Pendelausschlag in den Jahren 1960–1964 deutlich weniger gross als in den vorangehenden Jahren. Dieser Drift kann auch mit einer Exponentialfunktion nur annähernd erfasst werden.

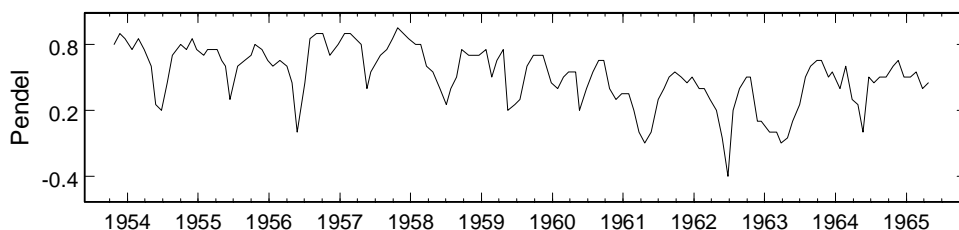


Bild 8-1: Pendel A, radial

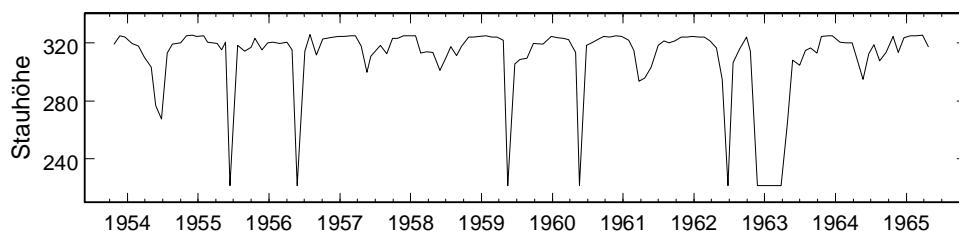


Bild 8-2: Stauhöhe

Von den 152 Beobachtungen wurden 15 gestrichen um einen einigermaßen gleichmässigen Zeitschritt zu bekommen. Trotzdem variieren die Zeitabstände noch zwischen 20

und 49 Tagen. Der Durbin-Watson-Test und andere Aussagen im Zusammenhang mit der Autokorrelation der Fehler müssen daher etwas vorsichtig interpretiert werden.

8.2 Langzeitprognose

Als erstes soll, wie in den vorhergehenden Beispielen, versucht werden anhand der ersten Jahre eine Prognose zu machen. Man kann aber hier nicht erwarten, dass die grosse Deformation ab 1960 vom Modell vorausgesagt wird. Leider ist für diese Studie nichts weiter bekannt über das aussergewöhnliche Verhalten der Mauer, und so kann auch keine vernünftige Interpretation gemacht werden.

Es werden die ersten 5 Jahre als Regressionsperiode gewählt. Das Ausgangsmodell umfasst 4 Tschebyscheff-Polynome, Sinus- und Kosinusfunktionen mit Perioden von einem ganzen und einem halben Jahr, und Exponentialfunktion e^{-t} , $e^{-t/2}$, $e^{-t/4}$, $e^{-t/8}$.

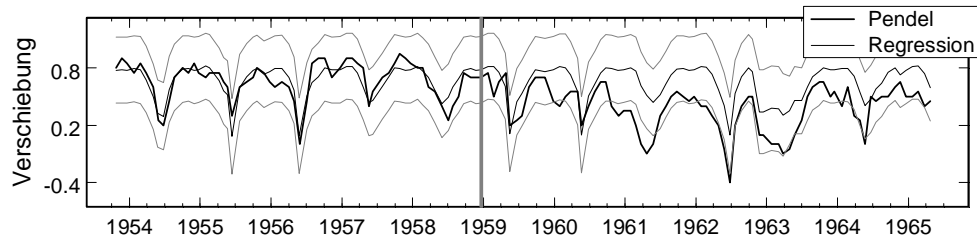
8.2.1 Klassische Methode

Mit der Methode der kleinsten Quadrate bekommt man zunächst sehr grosse VIF-Werte ($3 \cdot 10^5$), die auf die Exponentialfunktionen zurückzuführen sind. Eine Ridge-Regression mit einem Verzerrungsparameter $k = 0.05$ bringt den maximalen VIF-Wert hinunter auf 128. Die schrittweise Elimination führt dann zum Modell in Tabelle 8-1. Wie man sieht, sind die Exponentialfunktionen alle weggefallen. Das Modell sieht also keine Langzeitdeformation voraus.

Tabelle 8-1: Modell mit Prognose, $k = 0.005$

Regressor	Coefficient	Std Err	t-stat	p-value	VIF	Std Coeff
Constant	0.498894	0.03774	13.22	5.277E-19	0	0
T1(Cote)	0.220271	0.04531	4.862	9.52E-6	1.546	0.422163
sin(s)	0.0963774	0.02194	4.393	4.924E-5	1.218	0.338456
cos(s)	0.0855768	0.02092	4.09	0.0001369	1.192	0.311857
sin(2s)	-0.0695001	0.02095	-3.318	0.001585	1.132	-0.246473
SS Res	MS Res	R2	F	signif F	DW	Observ
0.672002	0.0117895	0.72205	36.79	3.484E-15	0.87	62
Regression: 23/10/1953-20/12/1958			Prediction: 24/01/1959-24/04/1965			
SS Res	MS' Res		SS Res	MS' Res		
0.672002	0.0108387		6.93126	0.0924168		

Das ist auch aus Bild 8-1 ersichtlich, wo der Vergleich zwischen Prognose und Messung dargestellt ist. Eingezeichnet sind auch die Grenzen für das 99.7% Prognoseintervall. Da auch in der Regression die Übereinstimmung ziemlich schlecht ist, wird dieses Intervall relativ breit und die gemessene Verschiebung überschreitet die Grenze nur zeitweise etwas.

Bild 8-3: Regression, Modell mit Prognose, $k = 0.005$, 99.7% Prognoseintervall

8.2.2 Modell mit hohen VIF-Werten

Es ist noch instruktiv zu sehen, was passieren kann, wenn eine starke Multikollinearität vorhanden ist. Im Modell in Tabelle 8-2 sind die Exponentialfunktionen wieder eingeführt. Die klassische Methode der kleinsten Quadrate führt zu hohen VIF-Werten. Die Exponentialfunktionen bilden eine Kombination, mit der sie die Regression etwas verbessern können. Jedoch ist die Prognose gänzlich unbrauchbar, wie Bild 8-4 zeigt. Das Gute ist, dass die Berechnung das Problem entschärft, indem das Prognoseintervall sich aufweitet und man sieht, dass etwas nicht stimmen kann. Ob man sich aber immer darauf verlassen kann, ist fraglich.

Interessant ist auch, dass das gleiche Modell mit einem Verzerrungsparameter $k = 0.005$ die VIF-Werte auf 128 reduziert und praktisch das gleiche Bild gibt wie Bild 8-3.

Tabelle 8-2: Modell mit Prognose und hohen VIF-Werten

Regressor	Coefficient	Std Err	t-stat	p-value	VIF	Std Coeff
Constant	-5.77819	1.905	-3.032	0.003723	0	0
T1(Cote)	0.223573	0.04291	5.21	3.034E-6	1.602	0.428492
sin(s)	0.0890217	0.0208	4.28	7.697E-5	1.264	0.312624
cos(s)	0.0900262	0.01971	4.568	2.895E-5	1.222	0.328071
sin(2s)	-0.0665606	0.01964	-3.389	0.001317	1.149	-0.236048
exp(-t/2)	6.23469	1.998	3.12	0.002902	1638	8.20475
exp(-t/4)	-22.452	7.004	-3.206	0.002264	1.327E4	-23.9894
exp(-t/8)	22.6009	6.954	3.25	0.00199	5745	16.0043
SS Res	MS Res	R2	F	signif F	DW	Observ
0.551119	0.0102059	0.77205	26.13	3.288E-15	1.07	62
Regression: 23/10/1953-20/12/1958			Prediction: 24/01/1959-24/04/1965			
SS Res	MS' Res	SS Res		MS' Res		
0.551119	0.00888902	77.0242		1.02699		

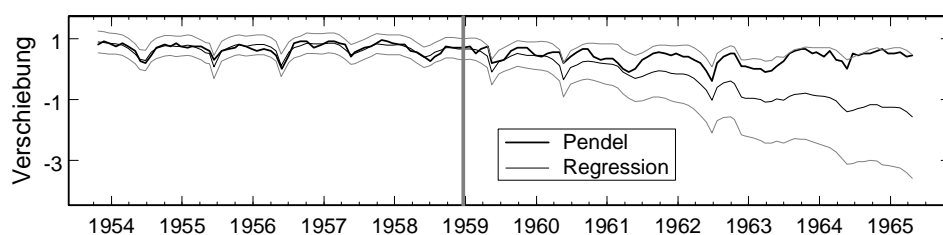


Bild 8-4: Regression, Modell mit Prognose und hohen VIF-Werten, 99.7% Prognoseintervall

8.3 Regression über ganze Messperiode

8.3.1 Modell ohne Langzeitfunktionen

Um das Bild abzurunden wird noch die Regression über die ganze Messperiode untersucht. In einem ersten Modell werden keine Exponentialfunktionen verwendet, damit man für die andern Fälle einen Vergleich hat. Nach der Elimination der Regressoren mit p -Werten über 1% bleibt das Modell in Tabelle 8-3. Das erste Polynom ist weggefallen und auch von den periodischen Funktionen ist nur die Kosinusfunktion noch ganz knapp im Modell geblieben. Dieses Beispiel zeigt, dass sich das Langzeitverhalten auch auf die andern Regressoren auswirken kann. Aus diesem Grunde sollte das Langzeitverhalten in der Regression modelliert werden, am besten mit Funktionen die sich nicht oder nur wenig auf die Prognose auswirken.

Die Residuen sind in Bild 8-5 geplottet. Die Grenzen des Prognoseintervalls liegen wegen der grossen Streuung ausserhalb der Figur.

Tabelle 8-3: Modell ohne Exponentialfunktionen

Regressor	Coefficient	Std Err	t-stat	p-value	VIF	Std Coeff
Constant	0.434762	0.02702	16.09	6.543E-33	0	0
T2(Cote)	0.124095	0.04443	2.793	0.006001	2.02	0.229754
T3(Cote)	0.264481	0.031	8.531	2.965E-14	1.858	0.673155
T4(Cote)	-0.220865	0.02868	-7.702	2.821E-12	1.793	-0.597008
cos(s)	0.0643931	0.02379	2.707	0.007691	1.25	0.175165
SS Res	MS Res	R2	F	signif F	DW	Observ
4.1817	0.0316795	0.55772	41.61	1.561E-22	0.54	137

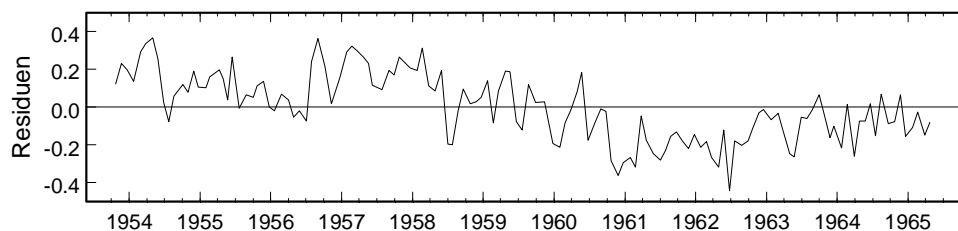


Bild 8-5: Residuen, Modell ohne Exponentialfunktionen

8.3.2 Modell mit Exponentialfunktionen

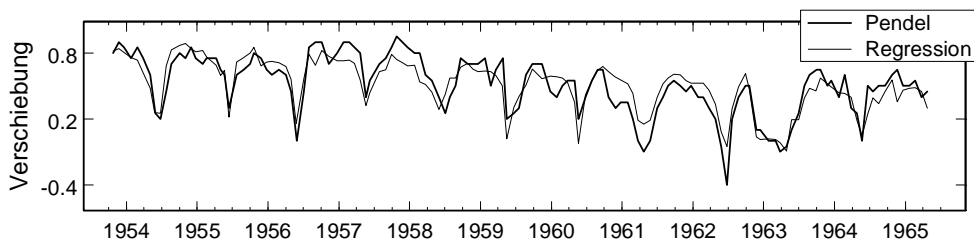
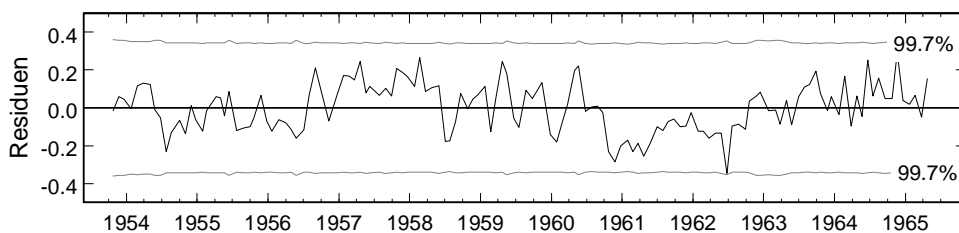
Beim nächsten Modell sind auch die Exponentialfunktionen dabei. Damit ergeben sich wieder hohe VIF-Werte (5000). Mit der Ridge-Regression mit $k = 0.005$ wird dieser Wert auf 118 hinunter gebracht. Die schrittweise Elimination führt zum Modell in Tabelle 8-4. Im Gegensatz zum letzten Modell sind sowohl das erste Polynom wie auch zwei Winkelfunktionen im Modell geblieben. Die Quadratsumme der Residuen hat sich fast halbiert.

Tabelle 8-4: Modell mit Exponentialfunktionen

Regressor	Coefficient	Std Err	t-stat	p-value	VIF	Std Coeff
Constant	0.0641199	0.03205	2	0.04754	0	0
T1(Cote)	0.253062	0.02348	10.78	9.524E-20	1.165	0.494661
T2(Cote)	0.115152	0.02486	4.632	8.679E-6	1.172	0.213199
cos(s)	0.0777747	0.01758	4.423	2.034E-5	1.266	0.211566
sin(2s)	-0.0504425	0.01636	-3.083	0.002507	1.066	-0.135325
exp(-t/2)	-0.559728	0.1676	-3.34	0.001092	13.34	-0.51878
exp(-t/4)	0.951069	0.1571	6.052	1.428E-8	13.36	0.940408
SS Res	MS Res	R2	F	signif F	DW	Observ
2.22149	0.0170884	0.76505	69.16	4.513E-38	0.85	137

Der hier gewählte Ansatz ist etwas verschieden von dem, der bei Antille (2001) benützt wurde. Die geschätzte Varianz MS_{Res} stimmt aber in etwa mit dem Quadrat der Standardabweichung (Antille, Tabelle 10.1) $\sigma^2 = 0.1376^2 = 0.01893$ überein. (Beim Wert von $R^2 = 0.9410$ in der gleichen Tabelle scheint es sich aber um einen Fehler zu handeln.)

Der Vergleich der Prognose mit der gemessenen Verschiebung ist in Bild 8-6 gezeigt. Die Residuen und das Korrelogramm sind in Bild 8-7 und Bild 8-8 geplottet.

**Bild 8-6: Regression, Modell mit Exponentialfunktionen****Bild 8-7: Residuen, Modell mit Exponentialfunktionen**

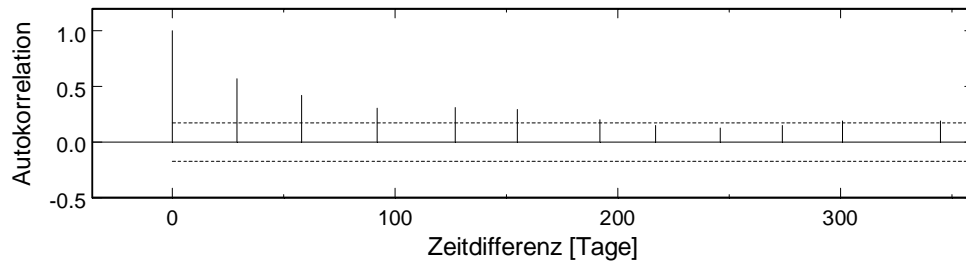


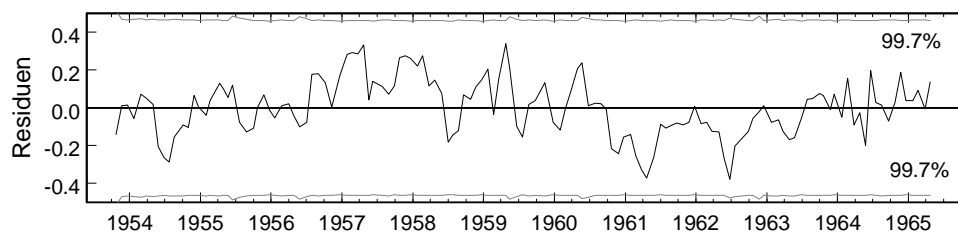
Bild 8-8: Korrelogramm, Modell mit Exponentialfunktionen

8.4 Autokorrelation der Fehler

Das Korrelogramm und der Durbin-Watson-Kennwert zeigen, dass auch bei diesem Beispiel die Fehler autokorreliert sind. Die schrittweise Elimination bis alle p -Werte unter 1% und alle VIF-Werte unter 10 liegen, führt zum Modell in Tabelle 8-5. Die Residuen sind etwas grösser als bei der klassischen Methode, wie das in der Regressionsperiode der Fall sein muss. Beim Vergleich müssen die Quadratsummen der Fehler des nicht transformierten Problems in der untersten Zeile der Tabelle betrachtet werden. Die Residuen und das Korrelogramm sind in Bild 8-9 und Bild 8-10 geplottet.

Tabelle 8-5: Modell mit Exponentialfunktionen, $\rho = 0.706$

Regressor	Coefficient	Std Err	t-stat	p-value	VIF	Std Coeff
Constant	0.236249	0.0502	4.706	6.283E-6	0	0
T1(Cote)	0.201513	0.02006	10.04	5.589E-18	1.08	0.562498
cos(s)	0.114286	0.02421	4.721	5.916E-6	1.038	0.25918
sin(2s)	-0.060846	0.01457	-4.175	5.367E-5	1.042	-0.229694
exp(-t/4)	0.415092	0.1119	3.708	0.0003064	1.167	0.215866
SS Res	MS Res	R2	F	signif F	DW	Observ
1.50674	0.0114147	0.61662	54	6.874E-27	2.00	137
Regression: 23/10/1953-24/04/1965						
SS Res	MS' Res					
3.00372	0.0219249					

Bild 8-9: Residuen, Modell mit Exponentialfunktionen, $\rho = 0.706$

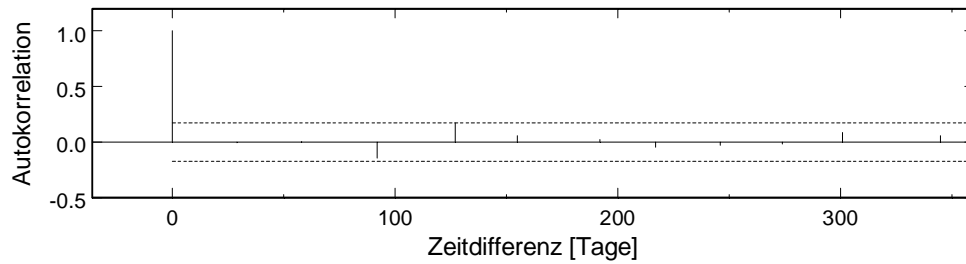


Bild 8-10: Korrelogramm, Modell mit Exponentialfunktionen, $\rho = 0.706$

8.5 Zusammenfassung Beispiel mit Langzeitverformungen

Das Beispiel mit Langzeitverformungen ist wegen der grossen Verformungen nach mehreren Jahren nicht unbedingt für eine Prognose geeignet. Die Langzeitverformungen können nicht vorausgesagt werden, und sollten es wahrscheinlich auch nicht. Ein Modell mit hohen VIF-Werten dient als Demonstration, wie sich eine Multikollinearität negativ auf die Prognose auswirken kann.

Neben der Prognose wurden auch Regressionen über die ganze Messperiode durchgeführt. Das Modell ohne Langzeitfunktionen zeigt, dass Langzeitabweichungen auch die Regression der übrigen Funktionen beeinträchtigen können. Das Modell mit Exponentialfunktionen über 2 und 4 Jahre zeigt eine befriedigende Übereinstimmung. Die Standardabweichung in der Semesterarbeit Antille (2001) konnte bestätigt werden.

Auch bei dieser Mauer sind die Fehler autokorreliert. Die verallgemeinerte Methode funktioniert gut, allerdings mit etwas grösseren Residuen, wie das für die Regression auch bei den früheren Beispielen beobachtet wurde.

9 Zusammenfassung und Schlussfolgerungen

Diese Zusammenfassung ist vor allem als Übersicht für den Anwender gedacht. Zusammenfassungen zu den Beispielen findet man in den jeweiligen Kapiteln. Die allgemeinen Schlussfolgerungen dieser Studie sind am Schluss dieses Kapitels festgehalten.

9.1 Modellierung und Einflussfunktionen

Am Anfang eines statistischen Modells steht die Modellierung des Verhaltens einer Staumauer. Da hier nur die lineare Regression betrachtet wird, müssen die gemessenen Daten oft noch transformiert werden, um den Einfluss richtig abbilden zu können. Hier werden die wichtigsten Modelle zusammengefasst, die bei Staumauern verwendet werden. Neben den klassischen werden auch weniger bekannte Einflussfunktionen einbezogen.

9.1.1 Stauhöhe

Der Einfluss der Stauhöhe kann im Allgemeinen gut mit einem Polynom bis vierten Grades modelliert werden. Bei einem rein statistischen Modell gilt dabei die Prognose nur für den Bereich der Stauhöhen, der für die Regression verwendet wurde. Falls Einflussfunktionen aus statischen Berechnungen vorhanden sind, können diese direkt verwendet werden (Spline-Funktionen oder Polynome) um diesen Bereich zu erweitern.

Damit die Regressoren möglichst unabhängig voneinander werden, sollte die Stauhöhe von einem Nullpunkt aus gemessen werden, der in der Nähe der gemessenen Daten liegt. Eine zusätzliche Skalierung ist dagegen nicht relevant. Nicht optimal aber akzeptabel ist der Minimalwert als Nullpunkt. Es entstehen grössere VIF-Werte, die sich aber in der Berechnung nicht negativ auswirken. Besser ist eine Zentrierung der Stauhöhe (Mittelwert Null) oder die Verwendung von Tschebyscheff-Polynomen. Damit werden die VIF-Werte klein gehalten, was die Beurteilung erleichtert, wenn andere Regressoren hohe VIF-Werte haben.

9.1.2 Periodische (saisonale) Effekte

Periodische Funktionen können benützt werden, wenn keine Temperaturmessungen vorhanden sind. Die klassischen Ansätze mit Sinus und Kosinus über ein Jahr und über ein halbes Jahr haben sich in der Praxis bewährt.

9.1.3 Temperatur

Betontemperaturen können in einem Regressionsmodell direkt verwendet werden. Physikalisch ist die Verformung einer Mauer vom gesamten Temperaturfeld in der Mauer abhängig. Messungen sind aber nur in bestimmten Punkten, meist nahe der Oberfläche vorhanden. Eine direkte Interpolation über die Mauerdicke entspricht nicht

den physikalischen Gesetzen der Wärmeleitung. Aus der Wärmeleitung ergibt sich, dass periodische Temperaturschwankungen im Innern der Mauer gegenüber jenen am Rand verkleinert, geglättet und verzögert auftreten. Die zeitliche Verzögerung kann für Jahresschwankungen mehrere Monate betragen. Um diese Effekte zu berücksichtigen, kann man mit Hilfe der eindimensionalen Wärmeleitung neue Temperatur-Regressoren generieren, die ungefähr den Temperaturen im Innern einer Mauer entsprechen. Die eindimensionale Wärmeleitung kann als Konvolutionsintegral formuliert werden, das auch bei unregelmässigen Zeitschritten numerisch einfach berechnet werden kann. Werden zu viele Temperatur-Regressoren generiert, führt das zu einer starken Multikollinearität, wodurch die Auswahl der Regressoren erschwert wird. Principal-Component-Regression und Ridge-Regression können dieses Problem nur bis zu einem gewissen Grad lösen. Deshalb empfiehlt es sich, nur wenige generierte Temperatur-Regressoren gezielt einzusetzen.

Lufttemperaturen können nicht direkt verwendet werden, da sie erst über die Betontemperaturen einen Einfluss auf die Verformung der Mauer haben. Mit Mittelwerten über Wochen und Monate kann aber eine Korrelation mit den Verschiebungen der Mauer gelingen. Auch raffiniertere Modelle, die die Wärmeübertragung von der Luft auf den Beton berücksichtigen, sind denkbar. Im Allgemeinen sind aber direkt gemessene Betontemperaturen vorzuziehen.

9.1.4 Kriechen

Ähnlich wie bei der Wärmeleitung kann aus der elastischen Verformung mit Hilfe eines Konvolutionsintegrals eine Kriechverformung bestimmt werden. Da man die elastische Verformung nicht kennt, muss für jeden Regressor der Stauhöhe ein Kriech-Regressor generiert werden. Die Kriecheffekte scheinen geringer zu sein als die Temperatureffekte.

9.1.5 Irreversible Effekte (Drift)

Irreversible Effekte werden meistens durch Exponentialfunktionen und lineare Funktionen modelliert. Ob man auch zunehmende Funktionen erlauben will, hängt von der Interpretation ab. Die Prognose sollte grundsätzlich nur das Normalverhalten ohne Drift beinhalten, damit eine Abweichung davon rechtzeitig festgestellt werden kann. Dies gilt insbesondere für ältere Mauern, bei denen keine Langzeitverformungen mehr zu erwarten sind. Hingegen kann man für eine Analyse auch zunehmende Funktionen einbeziehen, die dann eine eventuell vorhandene Langzeitverformung anzeigen. Vorsicht ist allerdings bei zunehmenden Exponentialfunktionen geboten.

Wenn Langzeiteffekte vorhanden sind, sollten sie in der Regression modelliert werden, da sonst die andern Regressoren verfälscht werden können. Damit die Prognose keine Langzeiteffekte beinhaltet, kann man Funktionen wählen, die wenig Einfluss auf die Prognose haben.

In dieser Studie wurden nur abnehmende Exponentialfunktionen benützt, nämlich $\exp(-t/T)$ mit $T = 1, 2, 4$ und 8 Jahren. Diese Funktionen sind leider stark voneinander abhängig, sodass die Auswahl wegen der Multikollinearität erschwert wird.

Eine beachtenswerte Variante, die hier nicht untersucht wurde, ist der Vorschlag von Fanelli et al. (2000). Die Langzeitverformungen werden durch Splinefunktionen modelliert. Diese sind sehr anpassungsfähig und haben keinen Einfluss auf die Prognose, da sie zeitlich lokal wirken. Mit dem so genannten Phasendiagramm kann der langfristige Trend trotzdem festgestellt werden.

9.2 Statistische Grössen

Hat man mit den oben beschriebenen Einflussfunktionen ein statistisches Modell aufgebaut, kann man mit der Methode der kleinsten Quadrate die Regressionskoeffizienten bestimmen. Neben den Koeffizienten werden aber noch viele andere statistische Grössen und Kennwerte berechnet, die es erlauben das gesamte Modell und auch die einzelnen Regressoren zu beurteilen. Anhand dieser Grössen kann man dann das Modell modifizieren bis man mit dem Resultat zufrieden ist. In diesem Abschnitt werden die einzelnen Grössen besprochen, die bei einer Berechnung auftreten. Strategien für die Modifikation, insbesondere für die Variablenauswahl, folgen später.

9.2.1 Beurteilungsgrössen für das gesamte Modell

Ein direktes Mass für die Qualität der Regression ist die **Quadratsumme der Residuen** SS_{Res} . Diese hängt jedoch von der Anzahl der Beobachtungswerte ab und ist physikalisch schwierig zu interpretieren. Besser in dieser Hinsicht ist die (geschätzte) **Standardabweichung** der Residuen $\hat{\sigma} = \sqrt{MS_{\text{Res}}}$. Sie ist mit der Anzahl der Freiheitsgrade normiert und hat die gleiche Masseinheit wie die Residuen selber.

Ein weiteres Mass für die Qualität einer Regression ist der Korrelationskoeffizient R oder dessen Quadrat, das **Bestimmtheitsmass** R^2 . Die Regression ist umso besser, je näher der Wert bei Eins liegt. Eine weitere Grösse ist die **globale F -Kenngrösse**. Eine grosse F -Kenngrösse gibt eine gute Korrelation an. Die F -Kenngrösse hat im Gegensatz zum Korrelationskoeffizienten eine statistische Bedeutung. Mit dem F -Test wird festgestellt ob mindestens ein Regressionskoeffizient signifikant von Null verschieden ist. Dies ist der Fall wenn die F -Kenngrösse grösser ist als ein kritischer Wert, oder wenn der zugehörige p -Wert kleiner ist als α , bei einem Signifikanzniveau α . Ein typischer Wert für α ist 1%.

Andere Grössen, wie das angepasste Bestimmtheitsmass, das auch die Anzahl der Regressoren berücksichtigt, weichen in allen Beispielen kaum vom gewöhnlichen Bestimmtheitsmass ab, da die Anzahl der Beobachtungen relativ gross ist.

9.2.2 Beurteilungsgrössen für einzelne Regressoren

Die Grössen die direkt verwendet werden, sind die **Regressionskoeffizienten**. Da sie aber nicht dimensionslos sind, sagt ihre Grösse nichts über ihren Einfluss auf die Regression. Etwas aussagekräftiger sind in dieser Hinsicht die **standardisierten Regressionskoeffizienten**. Das **Konfidenzintervall der Regressionskoeffizienten** gibt eine Abschätzung über die Streuung der Koeffizienten für das jeweilige Modell. Schliesst das Konfidenzintervall den Wert Null ein, so bedeutet dies, dass der Koeffizient gerade so gut als Null angenommen werden und damit weggelassen werden könnte.

Eine ähnliche Idee steckt hinter der ***t*-Kenngrösse**. Diese wird berechnet, indem der Regressionskoeffizient durch seinen Standardfehler (Schätzwert der Standardabweichung) dividiert wird. Sie gibt also an, wie weit der Regressionskoeffizient von Null entfernt ist, ausgedrückt als Vielfaches des Standardfehlers. Eine kleine *t*-Kenngrösse bedeutet also, dass der entsprechende Regressionskoeffizient nicht wesentlich von Null verschieden ist, wenn man die Streuung berücksichtigt. Der entsprechende Test ist der *t*-Test. Der ***p*-Wert** wird direkt aus der *t*-Kenngrösse bestimmt, wobei zusätzlich die Anzahl der Freiheitsgrade eine Rolle spielt. Bei einer grösseren Anzahl von Beobachtungsgrössen (ab etwa 50) ist aber dieser Einfluss vernachlässigbar. Eine grosse *t*-Kenngrösse entspricht einem kleinen *p*-Wert. Ein Regressor, der einen nichtvernachlässigbaren Einfluss hat, hat einen *p*-Wert von praktisch Null. Bei sehr kleinen *p*-Werten ist es einfacher die entsprechenden *t*-Kenngrössen zu vergleichen.

Die *t*-Kenngrösse hat eine zweite wichtige Interpretation. Sie ist ein Mass für die **Zunahme der Quadratsumme** der Residuen, wenn der entsprechende Regressor eliminiert wird. Ein Regressor mit einer kleinen *t*-Kenngrösse hat also auch in diesem Sinne eine geringere Bedeutung für das Modell.

9.2.3 Multikollinearität

Der **VIF** (Variance Inflation Factor) ist eine Grösse, die anzeigt, ob zwei oder mehrere Regressoren gegenseitig voneinander abhängen. Im Idealfall von unabhängigen Regressoren ist $VIF = 1$. Werte über 10 weisen auf eine starke Abhängigkeit hin. Man spricht dann auch von einer Multikollinearität. Der entsprechende Regressionskoeffizient kann nicht für sich allein interpretiert werden, sondern gilt nur für die entsprechende Kombination von Regressoren. Dies ist besonders von Bedeutung, wenn man den Regressionskoeffizienten direkt interpretieren will. Für die Prognose kann es wichtig sein, wenn nicht garantiert ist, dass die Regressoren ihre gegenseitige Abhängigkeit auch in der Zukunft beibehalten.

Hohe VIF-Werte und damit die Multikollinearität erschweren auch die Auswahl der Regressoren bei der Modellbildung. Wird ein Regressor eliminiert, ändern die übrigen ihren Koeffizienten und auch ihren *t*-Wert. Ein vorher bedeutender Regressor kann plötzlich unbedeutend werden und umgekehrt. Die Elimination sollte daher immer schrittweise erfolgen. Aber auch dann kann ein wichtiger Regressor frühzeitig aus dem Rennen fallen, obwohl er am Ende wichtig wäre.

9.2.4 Prognoseintervall und Extrapolation

Die Bandbreite der zulässigen Abweichung einer Beobachtung von der Prognose wird oft nach der Erfahrung definiert, z.B. als $\pm 3\hat{\sigma}$. Sie soll einerseits möglichst klein sein, damit ein anomales Verhalten möglichst bald erkannt wird, andererseits soll sie so gross sein, dass sie nicht zu oft überschritten wird. Da die Regressionskoeffizienten selber Zufallsvariablen sind, ist die Bandbreite theoretisch nicht konstant, sondern am kleinsten um den Mittelwert herum und grösser bei den Extremwerten der Regressoren. Die Statistik gibt eine Formel, mit der dieses variable Prognoseintervall bestimmt werden kann.

Bei der Prognose kann es wichtig sein, ob die neuen Beobachtungen im Bereich liegen, der bei der Regression verwendet wurde oder ob eine Extrapolation vorliegt. Dabei spielt nicht nur jeder einzelne Regressor für sich eine Rolle, sondern man muss ihre Kombination betrachten. Die Statistik gibt eine einfache Formel, mit der annähernd bestimmt werden kann, ob eine Extrapolation vorliegt.

9.2.5 Überprüfung der Annahmen

Die **Durbin-Watson-Kenngrösse** zeigt an, ob die Fehler autokorreliert sind. Ein Wert nahe von Null zeigt eine positive Autokorrelation an, während ein Wert nahe bei 2 bedeutet, dass keine Autokorrelation vorhanden ist. Sind die Fehler autokorreliert, ist eine wichtige Annahme der Methode der kleinsten Quadrate verletzt. Die Regressionskoeffizienten selber sind davon kaum betroffen, aber die t -Kenngrössen sind statistisch gesehen nicht mehr gültig. Sie sind aber numerisch immer noch ein Mass für die Zunahme der Residuen und können daher trotzdem noch hilfreich sein bei der Variablenauswahl. Der Durbin-Watson-Test überprüft nur eine bestimmte, allerdings häufig vorkommende Art der Autokorrelation der Fehler, nämlich diejenige, die durch einen autoregressiven Prozess erster Ordnung beschrieben werden kann.

Hilfreich bei der Beurteilung der Autokorrelation ist auch das **Korrelogramm**. Alle Werte ausser dem ersten bei der Zeitdifferenz Null sollten in einem Band von $\pm 2/\sqrt{n}$ liegen, wenn n die Anzahl der Residuen ist. Eine Autokorrelation zeigt sich dadurch, dass die Werte nur langsam abnehmen.

Die Residuen können in einem so genannten **Normalverteilungsplot** dargestellt werden. Liegen die Punkte auf einer Geraden, sind die Residuen normal verteilt.

9.3 Variablenauswahl

Die oben aufgeführten statistischen Grössen können helfen, aus einem Modell mit vielen Regressoren die wichtigsten auszulesen. Der direkteste Anhaltspunkt, welche Regressoren wichtig sind, gibt die t -Testgrösse und der davon abgeleitete p -Wert. Regressoren mit einem p -Wert über einem gewählten Grenzwert (typischerweise 1%) sind statistisch nicht signifikant und können eliminiert werden. Dabei ist zu beachten, dass dieses Kriterium nur für den jeweiligen Eliminationsschritt gilt, d.h. unter der Voraussetzung dass die andern Regressoren zum Modell gehören. Wird ein Regressor eliminiert, können die übrigen Regressoren ihre t -Kenngrössen ändern. Deshalb wird die schrittweise Elimination empfohlen.

Ist eine positive Autokorrelation der Fehler vorhanden, angezeigt durch eine kleine Durbin-Watson-Kenngrösse, so werden die t -Kenngrössen in der Regel überschätzt. D.h. nicht alle Regressoren mit einem kleinen p -Wert (unter 1%) sind auch wirklich statistisch signifikant. Haben einige Regressoren viel kleinere t -Kenngrössen als die übrigen, so ist das ein sicheres Zeichen, dass sie ebenfalls eliminiert werden sollten, auch wenn die entsprechenden p -Werte sehr klein sind. Das Problem tritt nicht mehr auf, wenn man die verallgemeinerte Methode der kleinsten Quadrate anwendet.

Bei einer Multikollinearität ist die gegenseitige Abhängigkeit der Regressoren besonders stark. Da in diesem Fall die t -Kenngrössen bei jedem Eliminationsschritt be-

sonders stark ändern, führt die schrittweise Elimination oft nicht zum optimalen Modell. Mann kann entweder verschiedene Varianten durchgehen oder zuerst die Multikollinearität beseitigen. Eine Methode, die oft zum Ziel führt, ist die Ridge-Regression.

Oft fallen Regressoren mit hohen VIF-Werten schon bei der Elimination aus dem Modell, da sie unwichtig sind. Manchmal kann man auch das Modell noch etwas weiter reduzieren und dadurch die Multikollinearität auflösen. Bleiben am Schluss noch Regressoren mit grossen VIF-Werten im Modell, kann man die Multikollinearität effektiv mit einer Principal-Component-Regression beseitigen. Oft ist es Ermessenssache, ob man versucht die VIF-Werte durch Elimination weiterer Regressoren oder durch eine Principal-Component-Regression zu reduzieren.

9.4 Weitergehende Methoden

In einigen Fällen, bei starken Multikollinearitäten oder wenn die Fehler autokorreliert sind, kommen aufwändigere Methoden zur Anwendung.

9.4.1 Principal-Component-Regression

Eine starke Multikollinearität kann durch die Principal-Component-Regression beseitigt werden. Es handelt sich um eine Eigenwertzerlegung, bei der Eigenwerte mit sehr kleinen Eigenwerten weggelassen werden.

Die Methode sollte nicht überbeansprucht werden. Enthält ein Modell sehr viele gegenseitig abhängige Regressoren, z.B. gemessene und berechnete Temperaturen, kann die Beseitigung der Multikollinearität schwierig werden. Es müssen dann sehr viele Eigenvektoren, oft mehr als die Hälfte, weggelassen werden bis die Multikollinearität verschwindet. Dadurch steigt die Quadratsumme der Residuen stark an, was bedeutet, dass die verbleibenden Eigenvektoren als Regressoren nicht mehr geeignet sind.

9.4.2 Ridge-Regression

Die Ridge-Regression ist eine einfache Alternative zur Principal-Component-Regression. Die Normalgleichungen werden durch einen Parameter k modifiziert, sodass die Kondition der Matrix verbessert wird. Der Parameter liegt im Bereich 0–1 und muss vom Anwender gewählt werden. In der Literatur wird empfohlen, dass man die Berechnung mit mehreren Werten für k durchführt und verfolgt, wie sich die Regressionsparameter verhalten.

Die Ridge-Regression ist nicht geeignet, eine sehr starke Multikollinearität zu beseitigen. Der Parameter muss in einem solchen Fall sehr gross gewählt werden, wodurch die Quadratsumme der Residuen übermässig ansteigt. Hingegen hat sich die Ridge-Regression für die Variablenauswahl in vielen Fällen bewährt. Schon ein kleiner Parameter (in den Beispielen 0.01 bis 0.001) kann die Kondition der Matrix soweit verbessern, dass die schrittweise Elimination aufgrund des t -Tests funktioniert. Die VIF-Werte müssen dazu nicht unter 10 sein, sondern können auch grösser sein. Da man mit verschiedenen Verzerrungsparametern oft zu verschiedenen Endmodellen kommt, sollte man immer mehrere Varianten ausprobieren.

9.4.3 Verallgemeinerte Methode der kleinsten Quadrate

Zeigt der Durbin-Watson-Kennwert eine Autokorrelation der Fehler an, so ist eine grundlegende Annahme der klassischen linearen Regression verletzt, nämlich die, dass die Fehler unkorreliert sind. Die verallgemeinerte Methode der kleinsten Quadrate kommt ohne diese Annahme aus und kann in diesem Fall eingesetzt werden. Der Durbin-Watson-Test zeigt nur eine bestimmte Art der Autokorrelation der Fehler an, nämlich diejenige, die durch einen autoregressiven Prozess erster Ordnung beschrieben werden kann. Für diesen Fall wird auch die verallgemeinerte Methode der kleinsten Quadrate relativ einfach. Man transformiert die Regressoren durch eine verallgemeinerte Differenzenbildung. Mit den transformierten Regressoren kann dann wieder die klassische Methode der kleinsten Quadrate verwendet werden. Eine Schwierigkeit ist die, dass für die Transformation der Autokorrelationsparameter ρ benötigt wird, der iterativ aus den Residuen berechnet werden kann. Allerdings ist die Konvergenz nicht garantiert, und wurde in einigen Beispielen auch nicht erreicht. In dieser Studie wurde nur der einfachste Algorithmus untersucht, aber es gibt auch aufwändigere Algorithmen, die auch andere Formen der Autokorrelation behandeln können.

Man kann sich fragen, was die Konsequenzen sind, wenn man trotz einer Autokorrelation der Fehler die klassische Methode anwendet. Die Annahme, dass die Fehler unkorreliert sind, wird nicht für die Berechnung der Regressionskoeffizienten benutzt. Diese werden daher richtig berechnet (unverzerrte Schätzwerte). Hingegen wird die Varianz der Koeffizienten, und damit die t -Kenngrösse falsch bestimmt. Man kann daher die Variablenauswahl nicht mehr rein nach dem t -Test durchführen. Meistens werden zu viele Regressoren als signifikant ausgewiesen. Allerdings gibt die t -Kenngrösse immer noch an, welche Regressoren die Residuen am wenigsten beeinflussen, und kann so immer noch für die Auswahl benutzt werden. Nur liegt die Anzahl der Regressoren, die eliminiert werden sollen, im Ermessen des Ingenieurs.

Die Theorie sagt auch, dass die Regressionskoeffizienten nicht mehr effektiv sind, d.h. nicht mehr minimale Varianz haben, wenn sie nicht mit der verallgemeinerten Methode berechnet werden. Man kann daher erwarten, dass die verallgemeinerte Methode eine bessere Prognose liefert, als wenn man die Autokorrelation der Fehler ignoriert. In den Beispielen hat sich das nur zum Teil bestätigt, und es ist auch nicht klar in welchen Fällen es zutrifft.

Insgesamt kann gesagt werden, dass die klassische Methode auch ihre Berechtigung hat, wenn die Annahmen nicht ganz erfüllt sind, solange man sich im Klaren ist, wie man die statistischen Grössen interpretieren muss. Andererseits scheint die verallgemeinerte Methode ein viel versprechender Ansatz für eine verbesserte Prognose zu sein. Hier müssten aber noch weitere Untersuchungen durchgeführt und mehr Erfahrung gewonnen werden.

9.5 Allgemeine Folgerungen

- Das klassische Modell mit einem Polynom der Stauhöhe und einem Sinus-Kosinus-Ansatz für saisonale Einflüsse hat sich als erstaunlich genau und robust erwiesen. Modelle mit gemessenen Temperaturen waren in den Beispielen kaum besser und Modelle mit berechneten Temperaturen nur wenig besser.
- Da nicht jedes Modell, das eine gute Übereinstimmung bei der Regression zeigt, auch eine gute Prognose liefert, empfiehlt es sich, verschiedene Intervalle für die Regression zu benützen und auch die Prognose mit gemessenen Werten zu überprüfen.
- Die ursprüngliche Idee, einfach möglichst viele Regressoren einzuführen und dann mit Hilfe der Statistik die wichtigsten auszuwählen, hat sich nicht bewährt, da in solchen Fällen oft eine starke Multikollinearität auftritt, die die Auswahl der Regressoren erschwert. Um dieses Problem zu vermeiden, sollte man keine unnötigen Multikollinearitäten einführen. Berechnete Grössen, wie Temperaturen im Innern der Mauer, sollten gezielt und physikalisch sinnvoll eingesetzt werden. In manchen Fällen ist es auch möglich orthogonale Regressoren zu konstruieren, wie z.B. durch die Verwendung von Tschebyscheff-Polynomen.
- Wenn die Fehler autokorreliert sind, kann die klassische Methode der kleinsten Quadrate immer noch zur Berechnung der Regressionskoeffizienten benützt werden, aber viele statistische Aussagen sind nicht mehr gültig. Trotz den theoretischen Problemen kann man aber praktisch immer noch eine gute Prognose erreichen. Man muss sich nur im Klaren sein, dass die t -Kenngrössen nicht mehr als absolutes Kriterium genommen werden dürfen.
- Im Falle einer Autokorrelation der Fehler bringt die verallgemeinerte Methode der kleinsten Quadrate gewisse Vorteile. Die t -Kenngrössen werden richtig berechnet, die Multikollinearität verringert sich, und in manchen Fällen wird eine bessere Prognose erzielt. Der einfachste Algorithmus, der in dieser Studie verwendet wurde, führt allerdings nicht immer zum Ziel. Eine weitere Untersuchung zu diesem Thema wäre nötig, bevor man die Methode allgemein einsetzen könnte.
- Obwohl sehr viele Berechnungen durchgeführt wurden, konnten nicht alle Fragen endgültig beantwortet werden. Eine weitergehende Studie, in der mehrere verschiedene Mauern untersucht werden, könnte hier weitere Klarheit bringen. Besonders für die Verfahren, die bisher nicht in der Praxis Anwendung gefunden haben, muss zuerst eine gewisse Erfahrung gewonnen werden, bevor man sie zur allgemeinen Anwendung empfehlen kann.

A Mathematische Herleitungen

A.1 Inverse von Blockmatrizen

In mehreren Herleitungen wird die Inverse einer Blockmatrix benötigt. Diese ist (Seber 1977, S. 390)

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^T & \mathbf{C} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{A}^{-1} + \mathbf{E}\mathbf{D}^{-1}\mathbf{E}^T & -\mathbf{E}\mathbf{D}^{-1} \\ -\mathbf{D}^{-1}\mathbf{E}^T & \mathbf{D}^{-1} \end{bmatrix}, \quad (\text{A.1})$$

wobei

$$\begin{aligned} \mathbf{D} &= \mathbf{C} - \mathbf{B}^T \mathbf{A}^{-1} \mathbf{B} \\ \mathbf{E} &= \mathbf{A}^{-1} \mathbf{B} \end{aligned} \quad (\text{A.2})$$

A.2 SS_{Res}

Die Quadratsumme der Residuen kann in Vektorform geschrieben werden als

$$\begin{aligned} SS_{\text{Res}} &= (\mathbf{y} - \hat{\mathbf{y}})^T (\mathbf{y} - \hat{\mathbf{y}}) \\ &= \mathbf{y}^T \mathbf{y} - \mathbf{y}^T \mathbf{X} \hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^T (\mathbf{X}^T \mathbf{y} - \mathbf{X}^T \mathbf{X} \hat{\boldsymbol{\beta}}) \\ &= \mathbf{y}^T (\mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}}). \end{aligned} \quad (\text{A.3})$$

A.3 VIF

Um zu zeigen, dass die Beziehung

$$\text{VIF} = \frac{1}{1 - R_j^2} = S_{jj} C_{jj} \quad (\text{A.4})$$

gilt, bildet man zunächst die Matrix

$$[\mathbf{X} \quad \mathbf{y}]^T [\mathbf{X} \quad \mathbf{y}] = \begin{bmatrix} \mathbf{X}^T \mathbf{X} & \mathbf{X}^T \mathbf{y} \\ \mathbf{y}^T \mathbf{X} & \mathbf{y}^T \mathbf{y} \end{bmatrix}. \quad (\text{A.5})$$

Der Diagonalterm der Inversen, der zu \mathbf{y} gehört, ist nach (A.1)

$$C_{yy} = \mathbf{D}^{-1} = \frac{1}{\mathbf{y}^T \mathbf{y} - \mathbf{y}^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}} = \frac{1}{\mathbf{y}^T (\mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}})} = \frac{1}{SS_{\text{Res}}}. \quad (\text{A.6})$$

Andererseits ist

$$-R^2 = \frac{SS_{\text{Res}}}{SS_{\text{Tot}}} = \frac{SS_{\text{Res}}}{S_{yy}}, \quad (\text{A.7})$$

oder

$$\frac{1}{1-R^2} = \frac{S_{yy}}{SS_{\text{Res}}} = S_{yy} C_{yy}. \quad (\text{A.8})$$

Betrachtet man anstatt dem Vektor \mathbf{y} eine Spalte \mathbf{x}_j aus der \mathbf{X} -Matrix, so sieht man sofort, dass die Beziehung (A.4) gilt. Dabei ist R_j der Korrelationskoeffizient, wenn man den Regressor \mathbf{x}_j als unabhängige Variable wählt und versucht durch die andern Regressoren auszudrücken. Ist diese Korrelation gut, so heisst das, dass der Regressor \mathbf{x}_j tatsächlich durch die andern Regressoren ausgedrückt werden kann, dass also eine Multikollinearität besteht.

A.4 Regressor eliminieren oder einfügen, partieller F-Test

A.4.1 Partiieller F-Test bei der Elimination

Die Kenngrösse F_0 beschreibt die Zunahme der Quadratsumme der Residuen, wenn ein Regressor eliminiert wird. Hier soll der Zusammenhang zwischen F_0 und der Kenngrösse t_0 gezeigt werden:

$$F_0 = \frac{SS_{\text{Res}}(\hat{\boldsymbol{\beta}}_H) - SS_{\text{Res}}(\hat{\boldsymbol{\beta}})}{MS_{\text{Res}}} = \frac{\hat{\beta}_j^2}{MS_{\text{Res}} C_{jj}} = t_0^2. \quad (\text{A.9})$$

$SS_{\text{Res}}(\hat{\boldsymbol{\beta}})$ und $SS_{\text{Res}}(\hat{\boldsymbol{\beta}}_H)$ sind die Quadratsummen der Residuen vor und nach der Elimination. Der Vektor $\hat{\boldsymbol{\beta}}$ enthält die Regressionskoeffizienten vor der Elimination, $\hat{\boldsymbol{\beta}}_H$ diejenigen nach der Elimination. Beide Vektoren haben die gleiche Dimension, wobei der Eintrag für den eliminierten Regressor in $\hat{\boldsymbol{\beta}}_H$ Null ist.

Um $\hat{\boldsymbol{\beta}}_H$ zu berechnen, kann man die Methode der kleinsten Quadrate mit der Nebenbedingung formulieren, dass ein Regressionskoeffizient Null sein muss. Man minimiert also

$$(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \quad (\text{A.10})$$

unter der Nebenbedingung

$$\mathbf{c}^T \boldsymbol{\beta} = 0. \quad (\text{A.11})$$

Um den Regressor β_j Null zu setzen, muss der Vektor \mathbf{c} wie folgt definiert werden:

$$\mathbf{c} = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \end{bmatrix} \begin{matrix} 0 \\ \\ j \\ \end{matrix} \quad (\text{A.12})$$

Die Nebenbedingung wird mit einem Lagrange-Multiplikator μ eingeführt. Dies führt zu einem Minimierungsproblem ohne Nebenbedingung für den folgenden Ausdruck:

$$(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + \mu \mathbf{c}^T \boldsymbol{\beta} = 0. \quad (\text{A.13})$$

Die Methode der kleinsten Quadrate führt zur Gleichung

$$-2\mathbf{X}^T \mathbf{y} + 2\mathbf{X}^T \mathbf{X} \hat{\boldsymbol{\beta}}_H + \mu \mathbf{c} = \mathbf{0} \quad (\text{A.14})$$

und zur Gleichung der Nebenbedingung (A.11). Der Vektor der Regressionskoeffizienten unter der Nebenbedingung wird

$$\begin{aligned}\hat{\beta}_H &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} - \frac{\mu}{2} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{c} \\ &= \hat{\beta} - \frac{\mu}{2} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{c}.\end{aligned}\quad (\text{A.15})$$

Der Lagrange-Multiplikator kann bestimmt werden, indem man $\hat{\beta}_H$ in die Nebenbedingung einsetzt:

$$\mathbf{c}^T \hat{\beta}_H = \mathbf{c}^T \hat{\beta} - \frac{\mu}{2} \mathbf{c}^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{c} = 0. \quad (\text{A.16})$$

Es folgt

$$\frac{\mu}{2} = \frac{\mathbf{c}^T \hat{\beta}}{\mathbf{c}^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{c}}. \quad (\text{A.17})$$

Damit werden die Regressionskoeffizienten $\hat{\beta}_H$ mit Nebenbedingung, ausgedrückt durch $\hat{\beta}$, jene ohne Nebenbedingung,:

$$\hat{\beta}_H = \hat{\beta} - \frac{\mathbf{c}^T \hat{\beta}}{\mathbf{c}^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{c}} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{c}. \quad (\text{A.18})$$

Die Quadratsumme der Residuen für die neuen Regressionskoeffizienten ist damit

$$\begin{aligned}SS_{\text{Res}}(\hat{\beta}_H) &= \mathbf{y}^T (\mathbf{y} - \mathbf{X} \hat{\beta}_H) \\ &= \mathbf{y}^T (\mathbf{y} - \mathbf{X} \hat{\beta}) + \frac{\mathbf{c}^T \hat{\beta}}{\mathbf{c}^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{c}} \mathbf{y}^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{c} \\ &= SS_{\text{Res}}(\hat{\beta}) + \frac{(\mathbf{c}^T \hat{\beta})(\hat{\beta}^T \mathbf{c})}{\mathbf{c}^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{c}}\end{aligned}\quad (\text{A.19})$$

Damit ist die Differenz der Quadratsummen

$$SS_{\text{Res}}(\hat{\beta}_H) - SS_{\text{Res}}(\hat{\beta}) = \frac{(\mathbf{c}^T \hat{\beta})^2}{\mathbf{c}^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{c}}. \quad (\text{A.20})$$

Mit der Definition (A.12) ist $\mathbf{c}^T \hat{\beta} = \hat{\beta}_j$ und $\mathbf{c}^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{c} = C_{jj}$ und die Differenz der Quadratsummen wird

$$SS_{\text{Res}}(\hat{\beta}_H) - SS_{\text{Res}}(\hat{\beta}) = \frac{\hat{\beta}_j^2}{C_{jj}}. \quad (\text{A.21})$$

Damit ist (A.9) gezeigt.

A.4.2 Partieller F-Test beim Einfügen

Beim Einfügen eines Regressors wird die Matrix \mathbf{X} um eine Spalte \mathbf{z} erweitert:

$$\begin{bmatrix} \mathbf{X}^T \\ \mathbf{z}^T \end{bmatrix} [\mathbf{X} \quad \mathbf{z}] = \begin{bmatrix} \mathbf{X}^T \mathbf{X} & \mathbf{X}^T \mathbf{z} \\ \mathbf{z}^T \mathbf{X} & \mathbf{z}^T \mathbf{z} \end{bmatrix}. \quad (\text{A.22})$$

Die Inverse ist:

$$\begin{bmatrix} \mathbf{X}^T \mathbf{X} & \mathbf{X}^T \mathbf{z} \\ \mathbf{z}^T \mathbf{X} & \mathbf{z}^T \mathbf{z} \end{bmatrix}^{-1} = \begin{bmatrix} \dots & \dots \\ -\frac{\mathbf{z}^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1}}{d} & \frac{1}{d} \end{bmatrix} \quad (\text{A.23})$$

mit

$$d = \mathbf{z}^T (\mathbf{z} - \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{z}). \quad (\text{A.24})$$

Die neuen Regressionskoeffizienten sind:

$$\begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\beta}_z \end{bmatrix} = \begin{bmatrix} \dots & \dots \\ -\frac{\mathbf{z}^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1}}{d} & \frac{1}{d} \end{bmatrix} \begin{bmatrix} \mathbf{X}^T \mathbf{y} \\ \mathbf{z}^T \mathbf{y} \end{bmatrix}. \quad (\text{A.25})$$

Es wird im Moment nur der eine neue Regressionskoeffizient $\hat{\beta}_z$ betrachtet. Er ist

$$\hat{\beta}_z = \frac{\mathbf{z}^T (\mathbf{y} - \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y})}{d} = \frac{\mathbf{z}^T (\mathbf{H} - \mathbf{I}) \mathbf{y}}{\mathbf{z}^T (\mathbf{H} - \mathbf{I}) \mathbf{z}} \quad (\text{A.26})$$

mit $\mathbf{H} = \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$. Die Abnahme der Quadratsumme der Residuen beim Einfügen von $\hat{\beta}_z$ ist die Zunahme beim Eliminieren, also nach (A.21)

$$\Delta SS_{\text{Res}} = d \hat{\beta}_z^2 = \frac{[\mathbf{z}^T (\mathbf{I} - \mathbf{H}) \mathbf{y}]^2}{\mathbf{z}^T (\mathbf{I} - \mathbf{H}) \mathbf{z}}. \quad (\text{1.27})$$

Mit den Residuen

$$\mathbf{e}_y = \mathbf{y} - \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} = (\mathbf{I} - \mathbf{H}) \mathbf{y} \quad (\text{1.28})$$

und

$$\mathbf{e}_z = \mathbf{z} - \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{z} = (\mathbf{I} - \mathbf{H}) \mathbf{z} \quad (\text{1.29})$$

kann die Abnahme auch geschrieben werden als

$$\Delta SS_{\text{Res}} = \frac{\mathbf{z}^T \mathbf{e}_y}{\mathbf{z}^T \mathbf{e}_z} = \frac{\mathbf{e}_z^T \mathbf{e}_y}{\mathbf{e}_z^T \mathbf{e}_z}. \quad (\text{A.30})$$

Dabei wurde benützt, dass $(\mathbf{H} - \mathbf{I})(\mathbf{H} - \mathbf{I}) = \mathbf{H} - \mathbf{I}$. Die F -Kenngrösse ist somit

$$F_0 = \frac{\Delta SS_{\text{Res}}}{(SS_{\text{Res}} - \Delta SS_{\text{Res}})/(n - k - 2)}. \quad (\text{A.31})$$

In obiger Gleichung beziehen sich SS_{Res} und k auf das ursprüngliche Modell vor dem Einfügen von β_z . Der Wert im Nenner ist die mittlere Quadratsumme bezogen auf das Modell nach dem Einfügen. Die Quadratsumme ist daher um ΔSS_{Res} verkleinert und die Anzahl der Freiheitsgrade um einen erhöht.

Die t -Testgrösse schliesslich lässt sich bestimmen als $t_0 = \pm \sqrt{F_0}$, wobei sich das Vorzeichen nach dem Vorzeichen von d richtet.

A.5 Standardisierte Variablen

Standardisierte Variablen sind folgendermassen definiert:

$$\tilde{\mathbf{x}}_j = \frac{\mathbf{x}_j - \mathbf{1}\bar{x}_j}{\sqrt{S_{jj}}} \quad \text{und} \quad \tilde{\mathbf{y}} = \frac{\mathbf{y} - \mathbf{1}\bar{y}}{\sqrt{S_{yy}}}, \quad (\text{A.32})$$

wobei der Vektor $\mathbf{1}$ ein Vektor der Dimension n mit lauter Einsen bedeutet. Die Variablen werden auf ihren Mittelwert zentriert und durch ihre Länge dividiert. Als erstes wird der Einfluss der Zentrierung untersucht.

A.5.1 Zentrierung

Die Mittelwerte können ausgedrückt werden als

$$\bar{x}_j = \frac{1}{n} \mathbf{1}^T \mathbf{x}_j \quad \text{und} \quad \bar{y} = \frac{1}{n} \mathbf{1}^T \mathbf{y}. \quad (\text{A.33})$$

Damit kann die Matrix mit den zentrierten Regressoren geschrieben werden als

$$\mathbf{X}_C = \mathbf{X}_R - \frac{1}{n} \mathbf{1}\mathbf{1}^T \mathbf{X}_R, \quad (\text{A.34})$$

wobei \mathbf{X}_R die Matrix \mathbf{X} ohne die Einsen in der ersten Spalte bezeichnet. Der Vektor mit den zentrierten y -Werten kann geschrieben werden als

$$\mathbf{y}_C = \mathbf{y} - \frac{1}{n} \mathbf{1}\mathbf{1}^T \mathbf{y}. \quad (\text{A.35})$$

Für die späteren Herleitungen sind folgende Zwischenresultate nützlich:

$$\begin{aligned} \mathbf{X}_C^T \mathbf{X}_C &= \left(\mathbf{X}_R^T - \frac{1}{n} \mathbf{X}_R^T \mathbf{1}\mathbf{1}^T \right) \left(\mathbf{X}_R - \frac{1}{n} \mathbf{1}\mathbf{1}^T \mathbf{X}_R \right) \\ &= \mathbf{X}_R^T \mathbf{X}_R - \frac{2}{n} \mathbf{X}_R^T \mathbf{1}\mathbf{1}^T \mathbf{X}_R + \frac{1}{n^2} \mathbf{X}_R^T \underbrace{\mathbf{1}\mathbf{1}^T \mathbf{1}\mathbf{1}^T}_n \mathbf{X}_R \\ &= \mathbf{X}_R^T \mathbf{X}_R - \frac{1}{n} \mathbf{X}_R^T \mathbf{1}\mathbf{1}^T \mathbf{X}_R \end{aligned} \quad (\text{A.36})$$

und

$$\begin{aligned} \mathbf{X}_C^T \mathbf{y}_C &= \left(\mathbf{X}_R^T - \frac{1}{n} \mathbf{X}_R^T \mathbf{1}\mathbf{1}^T \right) \left(\mathbf{y} - \frac{1}{n} \mathbf{1}\mathbf{1}^T \mathbf{y} \right) \\ &= \mathbf{X}_R^T \mathbf{y} - \frac{2}{n} \mathbf{X}_R^T \mathbf{1}\mathbf{1}^T \mathbf{y} + \frac{1}{n^2} \mathbf{X}_R^T \underbrace{\mathbf{1}\mathbf{1}^T \mathbf{1}\mathbf{1}^T}_n \mathbf{y} \\ &= \mathbf{X}_R^T \mathbf{y} - \frac{1}{n} \mathbf{X}_R^T \mathbf{1}\mathbf{1}^T \mathbf{y} = \mathbf{X}_C^T \mathbf{y} \end{aligned} \quad (\text{A.37})$$

Mit der Matrix

$$\mathbf{X}^T \mathbf{X} = \begin{bmatrix} \mathbf{1}^T \\ \mathbf{X}_R^T \end{bmatrix} \begin{bmatrix} \mathbf{1} & \mathbf{X}_R \end{bmatrix} = \begin{bmatrix} n & \mathbf{1}^T \mathbf{X}_R \\ \mathbf{X}_R^T \mathbf{1} & \mathbf{X}_R^T \mathbf{X}_R \end{bmatrix} \quad (\text{A.38})$$

werden die Normalgleichungen:

$$\begin{bmatrix} n & \mathbf{1}^T \mathbf{X}_R \\ \mathbf{X}_R^T \mathbf{1} & \mathbf{X}_R^T \mathbf{X}_R \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_R \end{bmatrix} = \begin{bmatrix} \mathbf{1}^T \mathbf{y} \\ \mathbf{X}_R^T \mathbf{y} \end{bmatrix}, \quad (\text{A.39})$$

wobei $\hat{\beta}_R$ den Vektor mit den Regressionskoeffizienten ohne $\hat{\beta}_0$ bezeichnet, also

$$\hat{\beta}_R = \begin{bmatrix} \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_k \end{bmatrix}. \quad (\text{A.40})$$

Die erste Gleichung in (A.39) wird nun durch n dividiert und mit $\mathbf{X}_R^T \mathbf{1}$ vormultipliziert:

$$\begin{bmatrix} \mathbf{X}_R^T \mathbf{1} & \frac{1}{n} \mathbf{X}_R^T \mathbf{1} \mathbf{1}^T \mathbf{X}_R \\ \mathbf{X}_R^T \mathbf{1} & \mathbf{X}_R^T \mathbf{X}_R \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_R \end{bmatrix} = \begin{bmatrix} \frac{1}{n} \mathbf{X}_R^T \mathbf{1} \mathbf{1}^T \mathbf{y} \\ \mathbf{X}_R^T \mathbf{y} \end{bmatrix}. \quad (\text{A.41})$$

Dann wird die erste von der zweiten Gleichung abgezählt, was zu folgender Beziehung führt:

$$\left(\mathbf{X}_R^T \mathbf{X}_R - \frac{1}{n} \mathbf{X}_R^T \mathbf{1} \mathbf{1}^T \mathbf{X}_R \right) \hat{\beta}_R = \left(\mathbf{X}_R^T - \frac{1}{n} \mathbf{X}_R^T \mathbf{1} \mathbf{1}^T \right) \mathbf{y}. \quad (\text{A.42})$$

Mit (A.36) und (A.37) kann dies auch geschrieben werden als

$$\mathbf{X}_C^T \mathbf{X}_C \hat{\beta}_R = \mathbf{X}_C^T \mathbf{y}_C. \quad (\text{A.43})$$

Man kann also die Normalgleichungen auch mit den zentrierten Daten aufstellen und dabei die Konstante weglassen. Mit andern Worten, die Regressionsebene (Hyperebene) geht durch den Schwerpunkt. Dies ist eine Verallgemeinerung der Beobachtung, die bereits bei der einfachen Regression gemacht wurde. Die Konstante $\hat{\beta}_0$ kann aus der ersten Gleichung in (A.39) berechnet werden als

$$\hat{\beta}_0 = \frac{1}{n} \mathbf{1}^T \mathbf{y} - \frac{1}{n} \mathbf{1}^T \mathbf{X}' \hat{\beta}_R = \bar{y} - \sum_{j=1}^k \hat{\beta}_j \bar{x}_j \quad (\text{A.44})$$

Für die t -Kenngrösse wird der Diagonalterm C_{jj} der inversen Matrix (A.38) benötigt. Mit der Formel für die inverse Blockmatrix (A.1) erhält man mit

$$\mathbf{D} = \mathbf{X}_R^T \mathbf{X}_R - \frac{1}{n} \mathbf{X}_R^T \mathbf{1} \mathbf{1}^T \mathbf{X}_R = \mathbf{X}_C^T \mathbf{X}_C \quad (\text{A.45})$$

und

$$\mathbf{E} = \frac{1}{n} \mathbf{1}^T \mathbf{X}_R \quad (\text{A.46})$$

die Inverse

$$\begin{bmatrix} n & \mathbf{1}^T \mathbf{X}_R \\ \mathbf{X}_R^T \mathbf{1} & \mathbf{X}_R^T \mathbf{X}_R \end{bmatrix}^{-1} = \begin{bmatrix} \frac{1}{n} + \frac{1}{n^2} \mathbf{1}^T \mathbf{X}_R (\mathbf{X}_C^T \mathbf{X}_C)^{-1} \mathbf{X}_R^T \mathbf{1} & -\frac{1}{n} \mathbf{1}^T \mathbf{X}_R (\mathbf{X}_C^T \mathbf{X}_C)^{-1} \\ -\frac{1}{n} (\mathbf{X}_C^T \mathbf{X}_C)^{-1} \mathbf{X}_R^T \mathbf{1} & (\mathbf{X}_C^T \mathbf{X}_C)^{-1} \end{bmatrix}. \quad (\text{A.47})$$

Bemerkenswert ist die Tatsache, dass die Untermatrix, die den Regressoren entspricht gerade die Inverse $(\mathbf{X}_C^T \mathbf{X}_C)^{-1}$ ist. Für die Diagonalterme heisst das, dass

$$(C_C)_{jj} = C_{jj}. \quad (\text{A.48})$$

Man kann also den Diagonalterm entweder aus der ursprünglichen Formulierung oder aus der Formulierung mit zentrierten Variablen berechnen.

A.5.2 Skalierung

In einem zweiten Schritt werden die Auswirkungen der Skalierung behandelt. Die Matrix $\tilde{\mathbf{X}}$ mit den standardisierten Regressoren kann geschrieben werden als

$$\tilde{\mathbf{X}} = \mathbf{X}_C \mathbf{S}^{-1}, \quad (\text{A.49})$$

wobei \mathbf{S} folgende Diagonalmatrix ist:

$$\mathbf{S} = \begin{bmatrix} \sqrt{S_{11}} & & \\ & \ddots & \\ & & \sqrt{S_{kk}} \end{bmatrix}. \quad (\text{A.50})$$

Gleichung (A.43) kann folgendermassen umgeformt werden:

$$\mathbf{S}^{-1} \mathbf{X}_C^T \mathbf{X}_C \mathbf{S}^{-1} \mathbf{S} \hat{\boldsymbol{\beta}}_R / \sqrt{S_{yy}} = \mathbf{S}^{-1} \mathbf{X}_C^T \mathbf{y}_C / \sqrt{S_{yy}}, \quad (\text{A.51})$$

oder

$$\tilde{\mathbf{X}}^T \tilde{\mathbf{X}} \hat{\mathbf{b}} = \tilde{\mathbf{X}}^T \tilde{\mathbf{y}}. \quad (\text{A.52})$$

Die standardisierten Regressionskoeffizienten sind also

$$\hat{\mathbf{b}} = \mathbf{S} \hat{\boldsymbol{\beta}}_R / \sqrt{S_{yy}}, \quad (\text{A.53})$$

oder einzeln

$$\hat{b}_j = \hat{\beta}_j \left(\frac{S_{jj}}{S_{yy}} \right)^{1/2} \quad \text{für } j = 1, \dots, k. \quad (\text{A.54})$$

Für den Diagonalterm C_{jj} folgt aus $(\tilde{\mathbf{X}}^T \tilde{\mathbf{X}})^{-1} = \mathbf{S} (\mathbf{X}_C^T \mathbf{X}_C)^{-1} \mathbf{S}$, dass

$$\tilde{C}_{jj} = S_{jj} (C_C)_{jj} = S_{jj} C_{jj}. \quad (\text{A.55})$$

Die Gleichung (A.47) erlaubt es auch, die inverse Matrix $(\mathbf{X}^T \mathbf{X})^{-1}$ aus der Inversen $(\tilde{\mathbf{X}}^T \tilde{\mathbf{X}})^{-1}$ mit standardisierten Variablen zu berechnen (Es muss zusätzlich noch die Skalierung berücksichtigt werden). Dies kann nützlich sein um eine einheitliche Implementierung mit der ursprünglichen und mit der standardisierten Formulierung zu erreichen. Ausser für die Berechnung der Korrelationskoeffizienten wird diese Matrix noch für einige andere Grössen benötigt, wie für das Prognoseintervall und die Bestimmung

der Extrapolation. Diese Grössen sind in der Literatur nur für die ursprüngliche Form angegeben.

Die Quadratsumme der Residuen kann mit standardisierten Variablen wie folgt geschrieben werden:

$$\begin{aligned}
 SS_{\text{Res}} &= (\mathbf{y} - \hat{\mathbf{y}})^T (\mathbf{y} - \hat{\mathbf{y}}) \\
 &= S_{yy} \left[\frac{(\mathbf{y} - \bar{y}) - (\hat{\mathbf{y}} - \bar{y})}{\sqrt{S_{yy}}} \right]^T \left[\frac{(\mathbf{y} - \bar{y}) - (\hat{\mathbf{y}} - \bar{y})}{\sqrt{S_{yy}}} \right] \\
 &= S_{yy} (\tilde{\mathbf{y}} - \tilde{\mathbf{X}}\hat{\mathbf{b}})^T (\tilde{\mathbf{y}} - \tilde{\mathbf{X}}\hat{\mathbf{b}}). \tag{A.56}
 \end{aligned}$$

Wie für die originalen Variablen kann dieser Ausdruck noch vereinfacht werden, indem die Normalgleichungen benützt werden:

$$\begin{aligned}
 SS_{\text{Res}} &= S_{yy} \left[\tilde{\mathbf{y}}^T \tilde{\mathbf{y}} - \tilde{\mathbf{y}}^T \tilde{\mathbf{X}}\hat{\mathbf{b}} - \hat{\mathbf{b}}^T (\tilde{\mathbf{X}}^T \tilde{\mathbf{y}} - \tilde{\mathbf{X}}^T \tilde{\mathbf{X}}\hat{\mathbf{b}}) \right] \\
 &= S_{yy} (\tilde{\mathbf{y}}^T \tilde{\mathbf{y}} - \tilde{\mathbf{y}}^T \tilde{\mathbf{X}}\hat{\mathbf{b}}). \tag{A.57}
 \end{aligned}$$

A.6 Partieller F-Test für Principal-Component-Regression

Bei der Principal-Component-Regression kann man zwei Arten von partiellem F -Test unterscheiden. Der eine sagt aus wie viel die Quadratsumme der Residuen zunimmt wenn ein Eigenvektor weggelassen wird, die andere wenn ein Regressor eliminiert wird. Der erste Fall ist der einfachere:

$$\begin{aligned}
 SS_{\text{Res}} &= S_{yy} (\tilde{\mathbf{y}}^T \tilde{\mathbf{y}} - \tilde{\mathbf{y}}^T \tilde{\mathbf{X}}\hat{\mathbf{b}}) \\
 &= S_{yy} \tilde{\mathbf{y}}^T \tilde{\mathbf{y}} - S_{yy} \tilde{\mathbf{y}}^T \tilde{\mathbf{X}} \mathbf{V}_1 \mathbf{\Lambda}_1^{-1} \mathbf{V}_1^T \tilde{\mathbf{X}}^T \tilde{\mathbf{y}}. \tag{A.58}
 \end{aligned}$$

Der erste Teil ist unabhängig von den Regressoren. Beim zweiten Teil ist der Beitrag eines Eigenvektors

$$\Delta SS_{\text{Res}} = S_{yy} \frac{(\mathbf{v}_i^T \tilde{\mathbf{X}}^T \tilde{\mathbf{y}})^2}{\lambda_i}. \tag{A.59}$$

Dies ist die Formel, die in Kapitel 4 verwendet wurde.

Für den zweiten Fall kann die Differenz der Quadratsummen analog zur Standard-Regression mit Hilfe eines Lagrange-Multiplikators bestimmt werden. Die Nebenbedingung für die ursprünglichen standardisierten Regressionskoeffizienten $\hat{\mathbf{b}}_H$,

$$\mathbf{c}^T \hat{\mathbf{b}}_H = 0, \tag{A.60}$$

wird zur Nebenbedingung für die transformierten Koeffizienten $\hat{\boldsymbol{\alpha}}_H$,

$$\mathbf{c}^T \hat{\mathbf{b}}_H = \mathbf{c}^T \mathbf{V}_1 \hat{\boldsymbol{\alpha}}_H = \mathbf{w}^T \hat{\boldsymbol{\alpha}}_H = 0. \tag{A.61}$$

Analog zu (A.15) wird

$$\begin{aligned}
\hat{\boldsymbol{\alpha}}_H &= \Lambda_1^{-1} \mathbf{V}_1^T \tilde{\mathbf{X}}^T \tilde{\mathbf{y}} - \frac{\mu}{2} \Lambda_1^{-1} \mathbf{w} \\
&= \hat{\boldsymbol{\alpha}} - \frac{\mu}{2} \Lambda_1^{-1} \mathbf{w}
\end{aligned} \tag{A.62}$$

Den Lagrange-Multiplikator bestimmt man wiederum indem man in die Nebenbedingung einsetzt:

$$\mathbf{w}^T \hat{\boldsymbol{\alpha}}_H = \mathbf{w}^T \hat{\boldsymbol{\alpha}} - \frac{\mu}{2} \mathbf{w}^T \Lambda_1^{-1} \mathbf{w} = 0, \tag{A.63}$$

also

$$\frac{\mu}{2} = \frac{\mathbf{w}^T \hat{\boldsymbol{\alpha}}}{\mathbf{w}^T \Lambda_1^{-1} \mathbf{w}}. \tag{A.64}$$

Die Quadratsumme der Residuen ist

$$\begin{aligned}
SS_{\text{Res}}(\hat{\boldsymbol{\alpha}}_H) &= S_{yy} \tilde{\mathbf{y}}^T (\tilde{\mathbf{y}} - \tilde{\mathbf{X}} \mathbf{V}_1 \hat{\boldsymbol{\alpha}}_H) \\
&= S_{yy} \tilde{\mathbf{y}}^T (\tilde{\mathbf{y}} - \tilde{\mathbf{X}} \hat{\boldsymbol{\beta}}) + S_{yy} \frac{\mathbf{w}^T \hat{\boldsymbol{\alpha}}}{\mathbf{w}^T \Lambda_1^{-1} \mathbf{w}} \tilde{\mathbf{y}}^T \tilde{\mathbf{X}} \mathbf{V}_1 \Lambda_1^{-1} \mathbf{w} \\
&= SS_{\text{Res}}(\hat{\boldsymbol{\alpha}}) + S_{yy} \frac{(\mathbf{w}^T \hat{\boldsymbol{\alpha}})(\hat{\boldsymbol{\alpha}}^T \mathbf{w})}{\mathbf{w}^T \Lambda_1^{-1} \mathbf{w}}
\end{aligned} \tag{A.65}$$

Damit ist die Differenz der Quadratsummen

$$SS_{\text{Res}}(\hat{\boldsymbol{\alpha}}_H) - SS_{\text{Res}}(\hat{\boldsymbol{\alpha}}) = S_{yy} \frac{(\mathbf{w}^T \hat{\boldsymbol{\alpha}})^2}{\mathbf{w}^T \Lambda_1^{-1} \mathbf{w}}. \tag{A.66}$$

Mit $\mathbf{w}^T \hat{\boldsymbol{\alpha}} = \mathbf{c}^T \hat{\mathbf{b}} = \hat{b}_j$ und mit $\mathbf{w} = \mathbf{V}_1^T \mathbf{c}$ bekommt man schliesslich

$$SS_{\text{Res}}(\hat{\boldsymbol{\alpha}}_H) - SS_{\text{Res}}(\hat{\boldsymbol{\alpha}}) = \frac{\hat{b}_j^2 S_{yy}}{\mathbf{c}^T \mathbf{V}_1 \Lambda_1^{-1} \mathbf{V}_1^T \mathbf{c}} = \frac{\hat{b}_j^2 S_{yy}}{(\mathbf{V}_1 \Lambda_1^{-1} \mathbf{V}_1^T)_{jj}}. \tag{A.67}$$

Dies entspricht genau der in Kapitel 4 verwendeten Formel.

Nebenbei sei bemerkt, dass man (A.66) auch für das Weglassen eines Eigenvektors verwenden kann. In diesem Falle muss man in \mathbf{w} für den wegzulassenden Eigenvektor 1 und den Rest 0 setzen. Man bekommt $SS_{\text{Res}}(\hat{\boldsymbol{\alpha}}_H) - SS_{\text{Res}}(\hat{\boldsymbol{\alpha}}) = S_{yy} \hat{\alpha}_i^2 \lambda_i$, und mit $\hat{\alpha}_i = \mathbf{v}_i^T \tilde{\mathbf{X}}^T \tilde{\mathbf{y}} / \lambda_i$ wieder (A.59).

A.7 Numerische Behandlung der Methode der kleinsten Quadrate

Zur numerischen Behandlung der Methode der kleinsten Quadrate gibt es mehrere Varianten, nämlich

- Direkte Lösung
- QR-Zerlegung
- Singulärwertzerlegung

Sie unterscheiden sich in erster Linie dadurch, wie sie sich verhalten, wenn die Matrix $\mathbf{X}^T \mathbf{X}$ fast singulär ist und in der Komplexität der Berechnungen. Eine detaillierte Ab-

handlung ist in Golub und Van Loan (1996) zu finden. Die Algorithmen für die QR-Zerlegung und die Singulärwertzerlegung stehen in der Programmbibliothek LAPACK (1999) zur Verfügung.

Die direkte Lösung wird hier nicht weiter untersucht. Die QR-Zerlegung und die Singulärwertzerlegung basieren beide auf einer orthonormalen Transformation von $\tilde{\mathbf{X}}$. Diese Transformationen ändern die 2-Norm nicht. Die 2-Norm eines Vektors ist definiert als $\|\mathbf{x}\|_2 = \sqrt{\mathbf{x}^T \mathbf{x}}$ und wird hier für die Formulierung der Residuen benützt.

A.7.1 Formulierung mit QR-Zerlegung

Jede Matrix $\tilde{\mathbf{X}}$ lässt die so genannte QR-Zerlegung zu. Bei Rechteckmatrizen mit mehr Zeilen als Spalten kann man die QR-Zerlegung folgendermassen schreiben:

$$\tilde{\mathbf{X}} = \mathbf{Q}\mathbf{R} = [\mathbf{Q}_1 \quad \mathbf{Q}_2] \begin{bmatrix} \mathbf{R} \\ \mathbf{0} \end{bmatrix} = \mathbf{Q}_1 \mathbf{R}. \quad (\text{A.68})$$

Die Matrix \mathbf{Q} ist orthonormal, d.h. $\mathbf{Q}^T \mathbf{Q} = \mathbf{Q}\mathbf{Q}^T = \mathbf{I}$, und \mathbf{R} ist eine obere Dreiecksmatrix. Bei der Methode der kleinsten Quadrate soll

$$\|\tilde{\mathbf{y}} - \tilde{\mathbf{X}}\mathbf{b}\|_2^2 \quad (\text{A.69})$$

minimiert werden. Setzt man $\tilde{\mathbf{X}} = \mathbf{Q}\mathbf{R}$ und multipliziert von links mit \mathbf{Q}^T , erhält man

$$\|\tilde{\mathbf{y}} - \mathbf{Q}\mathbf{R}\mathbf{b}\|_2^2 = \|\mathbf{Q}^T \tilde{\mathbf{y}} - \mathbf{Q}^T \mathbf{Q}\mathbf{R}\mathbf{b}\|_2^2 = \left\| \begin{bmatrix} \mathbf{Q}_1^T \tilde{\mathbf{y}} - \mathbf{R}\mathbf{b} \\ \mathbf{Q}_2^T \tilde{\mathbf{y}} \end{bmatrix} \right\|_2^2. \quad (\text{A.70})$$

Hier wurde die Tatsache benützt, dass die 2-Norm durch die Multiplikation mit einer orthonormalen Matrix nicht verändert wird. Der Ausdruck wird minimal wenn

$$\mathbf{R}\mathbf{b} = \mathbf{Q}_1^T \tilde{\mathbf{y}}. \quad (\text{A.71})$$

Dieses Gleichungssystem lässt sich einfach lösen, da \mathbf{R} eine Dreiecksmatrix ist. Die Quadratsumme der Residuen lässt sich direkt oder als $\|\mathbf{Q}_1^T \tilde{\mathbf{y}}\|_2^2$ berechnen.

Die QR-Zerlegung ist numerisch genauer als die direkte Lösung der Normalengleichungen. Allerdings ist der Rechenaufwand bei der Zerlegung auch grösser.

A.7.2 Singulärwertzerlegung

Jede Matrix $\tilde{\mathbf{X}}$ lässt die so genannte Singulärwertzerlegung zu:

$$\tilde{\mathbf{X}} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T \quad (\text{A.72})$$

Die Matrizen \mathbf{U} und \mathbf{V} sind orthonormal, d.h. $\mathbf{U}^T \mathbf{U} = \mathbf{U}\mathbf{U}^T = \mathbf{I}$ und analog für \mathbf{V} . Die Matrix $\mathbf{\Sigma}$ ist diagonal und enthält die so genannten Singulärwerte σ_j (nicht zu verwechseln mit der Varianz σ^2):

$$\Sigma = \begin{bmatrix} \sigma_1 & & & & \\ & \ddots & & & \\ & & \sigma_k & & \\ & & & \ddots & \\ & & & & 0 \end{bmatrix}. \quad (\text{A.73})$$

Die Singulärwerte sind positiv und nach absteigender Grösse geordnet. Ist die Matrix $\tilde{\mathbf{X}}$ singulär, sind die letzten Singulärwerte Null. Die Anzahl der von Null verschiedenen Singulärwerte ist der Rang der Matrix. Bei einer Rechteckmatrix der Dimension $n \times k$ mit $k < n$ ist der Rang höchstens k .

Bei der Methode der kleinsten Quadrate soll

$$\|\tilde{\mathbf{y}} - \tilde{\mathbf{X}}\mathbf{b}\|_2^2 \quad (\text{A.74})$$

minimiert werden. Man kann für $\tilde{\mathbf{X}}$ die Singulärwertzerlegung einsetzen und den Ausdruck mit \mathbf{U}^T vormultiplizieren. Da eine Multiplikation mit einer orthonormalen Matrix den Wert der Norm nicht ändert, erhält man:

$$\|\tilde{\mathbf{y}} - \mathbf{U}\Sigma\mathbf{V}^T\mathbf{b}\|_2^2 = \|\mathbf{U}^T\tilde{\mathbf{y}} - \Sigma\mathbf{V}^T\mathbf{b}\|_2^2 = \sum_{j=1}^k (\mathbf{u}_j^T \tilde{\mathbf{y}} - \sigma_j \alpha_j)^2, \quad (\text{A.75})$$

mit $\alpha_j = \mathbf{v}_j^T \mathbf{b}$. Die Summe von Quadratzahlen ist minimal, wenn jeder Summand minimal ist. Für die von Null verschiedenen Singulärwerte wird daher $\alpha_j = \mathbf{u}_j^T \tilde{\mathbf{y}} / \sigma_j$ gewählt, womit die entsprechenden Summanden Null sind. Für die Summanden mit $\sigma_j = 0$ ist der Beitrag unabhängig von α_j gleich $(\mathbf{u}_j^T \tilde{\mathbf{y}})^2$. Aus $\alpha_j = \mathbf{v}_j^T \hat{\mathbf{b}} = \mathbf{u}_j^T \tilde{\mathbf{y}} / \sigma_j$ erhält man die Lösung für die Regressionskoeffizienten

$$\hat{\mathbf{b}} = \sum_{j=1}^k \frac{\mathbf{u}_j^T \tilde{\mathbf{y}}}{\sigma_j} \mathbf{v}_j. \quad (\text{A.76})$$

Der Fehler ist

$$\|\tilde{\mathbf{y}} - \tilde{\mathbf{X}}\hat{\mathbf{b}}\|_2^2 = \sum_{j=k+1}^n (\mathbf{u}_j^T \tilde{\mathbf{y}})^2. \quad (\text{A.77})$$

A.7.3 Singulärwertzerlegung und Principal-Component-Regression

Die Singulärwertzerlegung ist eine Verallgemeinerung der Principal-Component-Regression. Betrachtet man

$$\tilde{\mathbf{X}}^T \tilde{\mathbf{X}} = \mathbf{V}\Sigma\mathbf{U}^T\mathbf{U}\Sigma\mathbf{V}^T = \mathbf{V}\Sigma^2\mathbf{V}^T = \mathbf{V}\Lambda\mathbf{V}^T, \quad (\text{A.78})$$

so sieht man, dass die Singulärwerte die Quadratwurzel aus den Eigenwerten sind, und \mathbf{V} die Eigenvektoren enthält. Die Bezeichnungen in Kapitel 4 wurden in Übereinstimmung mit dieser Interpretation gewählt.

Die Matrizen können entsprechend dem Rang der Matrix $\tilde{\mathbf{X}}$ in Untermatrizen aufgeteilt werden:

$$\mathbf{U} = \begin{bmatrix} \mathbf{U}_1 \\ \mathbf{U}_2 \end{bmatrix}, \quad \mathbf{V} = \begin{bmatrix} \mathbf{V}_1 \\ \mathbf{V}_2 \end{bmatrix} \quad \text{und} \quad \mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma}_1 & \\ & \mathbf{0} \end{bmatrix} \quad (\text{A.79})$$

Da die Blöcke mit Index 2 in der Zerlegung keinen Beitrag leisten, gilt auch

$$\tilde{\mathbf{X}} = \mathbf{U}_1 \mathbf{\Sigma}_1 \mathbf{V}_1^T. \quad (\text{A.80})$$

Die Matrix \mathbf{U}_1 kann aus

$$\mathbf{U}_1 = \tilde{\mathbf{X}} \mathbf{V}_1 \mathbf{\Sigma}_1^{-1} \quad (\text{A.81})$$

berechnet werden. Mit den reduzierten Matrizen \mathbf{U}_1 und \mathbf{V}_1 kann Gleichung (A.76) geschrieben werden als

$$\hat{\mathbf{b}} = \mathbf{V}_1 \mathbf{\Sigma}_1^{-1} \mathbf{U}_1^T \tilde{\mathbf{y}} = \mathbf{V}_1 \mathbf{\Sigma}_1^{-2} \mathbf{V}_1^T \tilde{\mathbf{X}}^T \tilde{\mathbf{y}}. \quad (\text{A.82})$$

Dies ist genau die Lösung, die schon bei der Principal-Component-Regression gefunden wurde.

Gleichung (A.77) beschreibt den Fehler für die standardisierten $\tilde{\mathbf{y}}$. Für die Quadratsumme der Residuen muss dieser Fehler mit dem Skalierungsfaktor multipliziert werden:

$$SS_{\text{Res}} = S_{yy} \tilde{\mathbf{y}}^T \mathbf{U}_2 \mathbf{U}_2^T \tilde{\mathbf{y}}. \quad (\text{A.83})$$

Um diesen Ausdruck mit der Principal-Component-Regression zu vergleichen, benützt man

$$\mathbf{U} \mathbf{U}^T = \mathbf{U}_1 \mathbf{U}_1^T + \mathbf{U}_2 \mathbf{U}_2^T = \mathbf{I}, \quad (\text{A.84})$$

und somit

$$\mathbf{U}_2 \mathbf{U}_2^T = \mathbf{I} - \mathbf{U}_1 \mathbf{U}_1^T = \mathbf{I} - \tilde{\mathbf{X}} \mathbf{V}_1 \mathbf{\Sigma}_1^{-2} \mathbf{V}_1^T \tilde{\mathbf{X}}^T. \quad (\text{A.85})$$

Damit ergibt sich

$$SS_{\text{Res}} = S_{yy} \tilde{\mathbf{y}}^T (\mathbf{I} - \tilde{\mathbf{X}} \mathbf{V}_1 \mathbf{\Sigma}_1^{-2} \mathbf{V}_1^T \tilde{\mathbf{X}}^T) \tilde{\mathbf{y}}, \quad (\text{A.86})$$

also wieder die ursprüngliche Formel mit $(\tilde{\mathbf{X}}^T \tilde{\mathbf{X}})^{-1}$ ersetzt durch $\mathbf{V}_1 \mathbf{\Sigma}_1^{-2} \mathbf{V}_1^T$.

Die Berechnung der Regression mit der Singulärwert-Zerlegung ist numerisch genauer als diejenige mit der QR-Zerlegung, und diese wieder ist genauer als die direkte Lösung der Normalgleichungen. Allerdings ist der Rechenaufwand auch bei der Singulärwert-Zerlegung am grössten.

A.8 Verallgemeinerte Methode der kleinsten Quadrate

A.8.1 Allgemeine Theorie

Die klassische Methode der kleinsten Quadrate nimmt an, dass die Fehler nicht autokorreliert sind und konstante Varianz haben. Das kann man in Matrixnotation schreiben als

$$E(\boldsymbol{\varepsilon} \boldsymbol{\varepsilon}^T) = \sigma^2 \mathbf{I}. \quad (\text{A.87})$$

Die generalisierte Methode der kleinsten Quadrate erlaubt eine allgemeinere Form der Fehler, nämlich

$$E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T) = \sigma_u^2 \boldsymbol{\Omega}. \quad (\text{A.88})$$

Die Strategie ist nun die, die Variablen so zu transformieren, dass die Annahmen der klassischen Methode der kleinsten Quadrate wieder erfüllt sind. Dazu sucht man eine Matrix \mathbf{P} die folgende Gleichung erfüllt:

$$\mathbf{P}\boldsymbol{\Omega}\mathbf{P}^T = \mathbf{I}. \quad (\text{A.89})$$

Das ursprüngliche Problem wird mit \mathbf{P} transformiert:

$$\mathbf{P}\mathbf{y} = \mathbf{P}\mathbf{X}\boldsymbol{\beta} + \mathbf{P}\boldsymbol{\varepsilon} \quad (\text{A.90})$$

oder

$$\mathbf{y}^* = \mathbf{X}^*\boldsymbol{\beta} + \mathbf{u}. \quad (\text{A.91})$$

Für dieses Problem ist tatsächlich

$$E(\mathbf{u}\mathbf{u}^T) = E(\mathbf{P}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T\mathbf{P}^T) = \sigma_u^2 \mathbf{P}\boldsymbol{\Omega}\mathbf{P}^T = \sigma_u^2 \mathbf{I}, \quad (\text{A.92})$$

d.h. die Annahmen der klassischen Methode der kleinsten Quadrate sind erfüllt.

Die Regressionskoeffizienten werden nach der Rückrechnung auf das ursprüngliche Problem

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^{*T}\mathbf{X}^*)^{-1}\mathbf{X}^{*T}\mathbf{y}^* = (\mathbf{X}^T\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\Omega}^{-1}\mathbf{y} \quad (\text{A.93})$$

und deren Varianz wird

$$\text{var}(\hat{\boldsymbol{\beta}}) = \sigma_u^2 (\mathbf{X}^{*T}\mathbf{X}^*)^{-1} = \sigma_u^2 (\mathbf{X}^T\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}. \quad (\text{A.94})$$

Die Varianz σ_u^2 kann geschätzt werden durch

$$\hat{\sigma}_u^2 = \frac{\mathbf{u}^T\mathbf{u}}{n-k-1} = \frac{\boldsymbol{\varepsilon}^T\boldsymbol{\Omega}^{-1}\boldsymbol{\varepsilon}}{n-k-1}, \quad (\text{A.95})$$

wobei $\boldsymbol{\varepsilon}$ die Residuen des ursprünglichen Problems sind.

Die Struktur der Matrix $\boldsymbol{\Omega}$ muss angenommen werden. Für den Spezialfall der Korrelation erster Ordnung ist

$$\boldsymbol{\Omega} = \begin{bmatrix} 1 & \rho & \rho^2 & \dots \\ \rho & 1 & \rho & \dots \\ \vdots & \vdots & \vdots & \vdots \\ \dots & \rho^2 & \rho & 1 \end{bmatrix}. \quad (\text{A.96})$$

Andere Fehlerstrukturen, wie Autokorrelation zweiter oder höherer Ordnung, können ebenfalls formuliert werden. Man beachte, dass die Struktur der Fehler angenommen werden muss und nicht aus den Residuen hergeleitet werden kann, da für n^2 Einträge in der Varianz-Kovarianz-Matrix nur n Residuen zur Verfügung stehen.

Die Matrix \mathbf{P} ist für den Spezialfall der Korrelation erster Ordnung ist

$$\mathbf{P} = \frac{1}{\sqrt{1-\rho^2}} \mathbf{T} \quad (\text{A.97})$$

mit der Transformationsmatrix

$$\mathbf{T} = \begin{bmatrix} \sqrt{1-\rho^2} & & & & \\ -\rho & 1 & & & \\ & -\rho & 1 & & \\ & & \ddots & \ddots & \\ & & & -\rho & 1 \end{bmatrix}. \quad (\text{A.98})$$

$\mathbf{T}\mathbf{y} = \mathbf{T}\mathbf{X}\boldsymbol{\beta} + \mathbf{T}\boldsymbol{\varepsilon}$ ist, wie man leicht sieht, die verallgemeinerte Differenzenbildung, wie sie in Kapitel 4 beschrieben wurde. Wie man verifizieren kann, gilt $\mathbf{T}\boldsymbol{\Omega}\mathbf{T}^T = (1-\rho^2)\mathbf{I}$. Der Erwartungswert für die Fehler des transformierten Systems ist damit

$$E(\mathbf{u}\mathbf{u}^T) = E(\mathbf{T}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T\mathbf{T}^T) = \sigma^2\mathbf{T}\boldsymbol{\Omega}\mathbf{T}^T = (1-\rho^2)\sigma^2\mathbf{I} = \sigma_u^2\mathbf{I} \quad (\text{A.99})$$

Damit ist auch die Beziehung zwischen der Varianz σ_u^2 der transformierten Fehler u_t und der Varianz σ^2 der Fehler ε_t gegeben (Montgomery, S. 490):

$$\text{var}(\varepsilon_t) = \sigma_u^2 \left(\frac{1}{1-\rho^2} \right). \quad (\text{A.100})$$

Alle diese verallgemeinerten Formeln und weitere Ableitungen werden nicht benötigt, wenn man direkt mit den transformierten Variablen arbeitet.

A.8.2 Schätzung des Autokorrelationsparameters

Nach Pesaran und Slater (1980, S. 37) kann eine Schätzung von ρ aus folgender Gleichung bestimmt werden:

$$-(n-1)a_2\hat{\rho}^3 - (n-2)a_3\hat{\rho}^3 + (na_2 + a_1)\hat{\rho} + na_3 = 0, \quad (\text{A.101})$$

mit

$$a_1 = \sum_{i=1}^n e_i^2, \quad a_2 = \sum_{i=2}^{n-1} e_i^2 \quad \text{und} \quad a_3 = \sum_{i=2}^n e_i e_{i-1}. \quad (\text{A.102})$$

Diese Formel kann über eine Maximum-Likelihood-Formulierung hergeleitet werden und sollte besser konvergieren als die einfache Formel in Kapitel 4. Die Gleichung (A.101) kann numerisch gelöst werden. Zu beachten ist, dass eventuell mehrere reelle Lösungen existieren.

Literatur

- Antille, Nicolas (2001), *Analyse des déformations des barrages à l'aide des séries temporelles*, Chaire de Statistique EPFL, projet de semestre, hiver 2000-2001.
- Carslaw, H.S., and J.C. Jaeger (1959), *Conduction of Heat in Solids*, 2nd. Ed., Oxford at the Clarendon Press.
- Chatterjee, Samprit and Bertram Price (1977), *Regression Analysis by Example*, John Wiley.
- Cochrane, D. and G.H. Orcutt (1949), "Application of Least Squares Regression to Relationships Containing Autocorrelated Error Terms", *Journal of the American Statistical Association*, Vol. 44, 32–61.
- Durbin, J. and G.S. Watson (1950), "Testing for Serial Correlation in Least Squares Regression. I", *Biometrika*, Vol. 37, 409–428.
- Durbin, J. and G.S. Watson (1951), "Testing for Serial Correlation in Least Squares Regression. II", *Biometrika*, Vol. 38, 159–178.
- Fanelli, M., G. Giusepetti, and G. Mazza (2000), "Analsis of the phenomenon of time drift in the observational data of dam behaviour", Paper R67 of Q78, 20th ICOLD Congress Beijing, p.1083.
- FERC (1999), *Engineering Guidelines for the Evaluation of Hydropower Projects*, Chapter 11–Arch Dams, Federal Energy Regulatory Commission, Division of Dam Safety and Inspections.
http://www.ferc.fed.us/hydro/docs/engguide/pdf_files/chap11.pdf
- Furnival, George M. and Robert W. Wilson (1974), "Regression by Leaps and Bounds", *Technometrics*, Vol 16. No. 4, 499–511.
- Goldberger, Arthur S. (1966), *Econometric Theory*, John Wiley.
- Golub, H. Gene and Charles F. Van Loan (1996), *Matrix Computations*, 3rd Ed., Johns Hopkins University Press.
- Harvey A.C. (1992), *Zeitreihenmodelle*, Oldenburg Verlag.
- ICOLD (2002), Sixth ICOLD Benchmark Workshop on Numerical Analysis of Dams, Salzburg 17-19 October.
- Kmenta, Jan (1997), *Elements of Econometrics*, 2nd Edition, The University of Michigan Press.
- LAPACK (1999), E. Anderson et al., *LAPACK Users' Guide*, 3rd Ed., Society for Industrial and Applied Mathematics.

- Mansfield, E. R., J.T. Webster and R. F. Gunst (1977), “An Analytic Variable Selection Technique for Principal Component Regression”, *Appl. Statist.*, Vol. 26, No. 1, 34–40.
- Montgomery, Douglas C., Elizabeth A. Peck, G. Geoffrey Vining (2001), *Introduction to Linear Regression Analysis*, Third Edition, John Wiley.
- Pesaran, M.H. and L.J. Slater (1980), *Dynamic Regression: Theory and Algorithms*, Halsted Press.
- Pindyck, Robert S and Daniel L. Rubinfeld (1998), *Econometric Models and Economic Forecasts*, Fourth Ed., McGraw-Hill.
- Schweizerisches Talsperrenkomitee (2000), Arbeitsgruppe “Numerik”, *Prédiction du Comportement des Barrages: Comparaison Mesuré-Calculé*, Entwurf Bericht.
- Seber, G. A .F. (1977), *Linear Regression Analysis*, John Wiley.