
Statistische Modellierung und Optimierung multipler Zielgrößen

Dissertation

zur Erlangung des Grades
eines Doktors der Naturwissenschaften
der Technischen Universität Dortmund

Der Fakultät Statistik
der Technischen Universität Dortmund

vorgelegt von

Nikolaus Rudak

Vorgelegt am: 21.01.2016

Tag der mündlichen Prüfung: 04.04.2016

1. Gutachter: Prof. Dr. Sonja Kuhnt

2. Gutachter: Prof. Dr. Claus Weihs

Für Marlon

Inhaltsverzeichnis

1. Einleitung	1
2. Statistische Modellierung eines thermischen Spritzprozesses	4
2.1. Thermische Spritzverfahren	4
2.2. Beschreibung des Versuchsaufbaus	7
2.3. Statistische Herausforderungen und bisherige Ergebnisse	8
3. Simultane Optimierung multipler Zielgrößen	15
3.1. Wünschbarkeitsbasierte Methoden	17
3.2. Verlustbasierte Methoden	20
3.3. Verlustbasierte JOP Methode	23
4. Simultane Modellierung des Erwartungswerts und der Varianz	26
4.1. Generalisierte lineare Modelle	26
4.2. Doppelt generalisierte lineare Modelle	30
5. Simultane Modellierung des Erwartungswertvektors und der Kovarianzmatrix	33
5.1. Multivariater Nadaraya-Watson-Kernschätzer	35
5.2. Schätzer basierend auf dem Matrix-Logarithmus-Modell mit einer allgemeinen Linkfunktion	37
5.3. Schätzer basierend auf dem Kovarianzregressionsmodell mit einer allgemeinen Linkfunktion	49
5.4. Heuristischer Ansatz	61
5.5. Simulationsstudie zum Vergleich der Methoden	62
6. Eigenschaften und Implementierung der JOP-Methode	79
6.1. Pareto-Optimalität der JOP-Methode	82
6.2. Algorithmus und Implementierung im R-Paket JOP	86

7. Erweiterung der JOP-Methode auf multiple korrelierte Zielgrößen	91
8. Anwendungsbeispiele	100
8.1. Wheel Cover Experiment	100
8.2. Thermisches Spritzexperiment	106
9. Zusammenfassung und Ausblick	115
A. Appendix	125
A.1. Newton-Algorithmus	125
A.2. JOP für unkorrelierte Zielgrößen: Liste der Eingabe- und Ausgabeargumente	126
A.3. Parameter für die Simulationsstudie in Kapitel 5.5	129
A.4. Datensätze	132
B. Notation	136

1. Einleitung

In der industriellen Praxis sind oft Maschineneinstellungen zu finden, die zu einem Produkt mit gewünschten Produkteigenschaften und mit möglichst geringer Unsicherheit führen, wie etwa bei einem thermischen Spritzprozess. In der Praxis werden Anwender mit Schwierigkeiten unterschiedlichster Art konfrontiert. Experimente können sehr zeitaufwändig und kostspielig sein, so dass nur eine geringe Anzahl von Versuchen möglich ist. Dies ist etwa beim thermischen Spritzprozess der Fall, welcher den Fokus im Teilprojekt B1 des SFB 823 darstellt. Hier wird insbesondere ein HVOF-Prozess (High Velocity Oxygen Fuel) betrachtet. Ziel ist es, Maschineneinstellungen zu finden, die zu einer gespritzten Schicht mit möglichst guten Eigenschaften (wie etwa Härte, Dicke, Porosität) führen. Der Prozess unterliegt tagesbedingten Schwankungen (Rehage et al., 2012b) und die Versuche sind sehr kostspielig und aufwändig. Aufgrund dessen werden die Versuche möglichst an einem Tag gefahren und daher kann ein Versuchsplan maximal 30 Versuche umfassen. Hinzu kommt, dass die Schichteigenschaften nur durch zerstörende aufwändige Prüfungen analysiert werden können, welche sich über Wochen ziehen können.

Zudem ist es oft der Fall, dass die gewünschten Produkteigenschaften nicht alle gleichzeitig erreicht werden können und der Anwender sich folglich mit einem möglichst guten Kompromiss zufrieden geben muss. Diese Problemstellung führt zu einem multikriteriellen Optimierungsproblem, welches mit verschiedenen, meistens regressionsbasierten, Methoden angegangen werden kann (vgl. z.B. Ardakani und Wulff (2013)). Sind die Zielgrößen zusätzlich korreliert, so sollte diese Information berücksichtigt werden, um eine möglichst realitätsnahe Lösung zu finden. Klassische Methoden, wie etwa die Verwendung von Konturplots (Myers et al., 2009) oder die bei Anwendern sehr beliebten Wünschbarkeitsfunktionen (Derringer und Suich, 1980), nutzen diese Informationen in der Regel nicht. Im einfachsten Fall mit sehr wenigen Einstellgrößen und Zielgrößen kann eine Lösung visuell mithilfe von Konturplots gesucht werden. Des Weiteren können abhängig von der Problemstellung die einzelnen Regressionsmodelle pro Zielgröße in Wünschbarkeitsfunktionen, die Werte zwischen Null und Eins annehmen, transformiert

werden. Anschließend wird das Produkt der einzelnen Wünschbarkeitsfunktionen maximiert. Bei dieser Methode müssen abhängig von der Problemstellung obere und untere Schranken pro Zielgröße spezifiziert werden.

Es existieren in der Literatur vielfach Erweiterungen der wünschbarkeitsbasierten Methode (Costa et al., 2011), die durch Verwendung spezieller Wünschbarkeitsfunktionen auch Korrelationen zwischen den Zielgrößen berücksichtigen (Wu, 2005, 2009). Andere wünschbarkeitsbasierte Methoden nutzen die Hauptkomponentenanalyse als Zwischenschritt (Wonggasem, 2012; Salmasnia et al., 2012, 2013b; Salmasnia und Bashiri, 2015). In Hamada und Balakrishnan (1998) werden empirische Mittelwerte, transformierte Varianzen und Korrelationen mithilfe von gewöhnlichen linearen Modellen modelliert und die Wahrscheinlichkeit, dass der Vektor der Zielgrößen innerhalb einer festgelegten Region liegt, wird auf Basis der multivariaten Normalverteilung maximiert. In Khuri und Conlon (1981) wird ein multivariates generalisiertes lineares Modell verwendet und eine multivariate Verlustfunktion minimiert. Eine weitere Klasse umfasst verlustbasierte Methoden (Pignatiello, 1993; Vining, 1998; Ko et al., 2005). Dabei wird der Erwartungswert einer multivariaten Verlustfunktion, die Risikofunktion, minimiert. Die Risikofunktion hängt hier von einem bedingten Erwartungswertvektor $E(Y|x)$ und einer bedingten Kovarianzmatrix $Cov(Y|x)$ ab, wobei Y für den Zielvariablenvektor und x für den Kovariablenvektor steht. Hier muss eine Kostenmatrix festgelegt werden. Meist wird eine einzige Kostenmatrix festgelegt und die Risikofunktion dafür minimiert. Die JOP-Methode (Kuhnt und Erdbrügge, 2004) schlägt die Minimierung der Risikofunktion für eine ganze Reihe von Kostenmatrizen vor. Dabei werden die Zielgrößen zunächst univariat mit doppelt generalisierten linearen Modellen (Smyth, 1989) abgebildet und anschließend wird die Risikofunktion für eine ganze Reihe von diagonalen Kostenmatrizen minimiert. Die Ergebnisse können dann grafisch mit Hilfe eines Joint Optimization Plots dargestellt und anhand dessen kann ein geeigneter Kompromiss ausgewählt werden.

In dieser Arbeit werden einige Eigenschaften der JOP-Methode präsentiert. Die klassische JOP-Methode berücksichtigt keine Korrelationen zwischen den Zielgrößen. Daher

wird die JOP-Methode in dieser Arbeit erweitert, um auch Korrelation zwischen den Zielgrößen aufzunehmen. Dies geschieht durch Verwendung von multivariaten Regressionsmodellen, bei denen sowohl der Erwartungswertvektor als auch die Kovarianzmatrix von Kovariablen abhängen darf. Dazu werden verschiedene Modelle (Hoff und Niu, 2012; Chiu et al., 1996; Yin et al., 2010) vorgestellt und erweitert. Des Weiteren werden zum Vergleich eine Heuristik sowie die Berechnung eines empirischen Mittelwerts und der empirischen Kovarianzmatrix hinzugezogen. Die Schätzung der unterschiedlichen Modelle wird in Simulationsstudien untersucht und verglichen. Im weiteren Verlauf wird eine mögliche Wahl von nichtdiagonalen Kostenmatrizen vorgeschlagen. Zudem wird ein Kriterium zur Bestimmung einer geeigneten Kompromisslösung präsentiert sowie die Pareto-Optimalität nachgewiesen.

Diese Arbeit ist wie folgt aufgebaut. In Kapitel 2 wird zunächst die Problemstellung anhand eines thermischen Spritzprozesses motiviert und die bisher verfassten Arbeiten zusammengefasst. Anschließend werden in Kapitel 3 verschiedene Methoden zur simultanen Optimierung multipler Zielgrößen vorgestellt. In Kapitel 4 werden generalisierte und doppelt generalisierte lineare Modelle präsentiert. Im Anschluss daran erfolgt in Kapitel 5 die Vorstellung und Erweiterung von multivariaten Modellen, bei denen sowohl der Erwartungswertvektor als auch die Kovarianzmatrix von Kovariablen abhängen darf. Das Ende des Kapitels bilden die Ergebnisse einer Simulationsstudie. In Kapitel 6 werden einige Eigenschaften der JOP-Methode dargestellt sowie das R-Paket JOP präsentiert. Die Erweiterung der JOP-Methode auf korrelierte Zielgrößen erfolgt in Kapitel 7. Verschiedene Anwendungsbeispiele werden in Kapitel 8 vorgestellt. Den Abschluss der Arbeit bildet die Zusammenfassung und der Ausblick in Kapitel 9.

2. Statistische Modellierung eines thermischen Spritzprozesses

Die in dieser Arbeit entwickelten statistischen Methoden werden durch eine Anwendung im Bereich des thermischen Spritzens motiviert. Objekte werden dabei mit einer Schicht versehen, welche wünschenswerte Eigenschaften besitzen soll. Beeinflusst wird die Qualität der Schicht von einstellbaren Anlageparametern an der Maschine sowie von unkontrollierbaren Einflüssen, wie etwa der Umgebungstemperatur. Bisher wurden zur statistischen Modellierung der Schichteigenschaften lediglich lineare Modelle mit Anlageparametern als Regressoren verwendet (Tillmann et al., 2010; Saaedi et al., 2010; Ramachandran et al., 2011; Hasan und Stokes, 2011; Kaßner et al., 2011; Gao et al., 2012). Um auch tagesabhängige Schwankungen im Prozess mit aufzunehmen, werden weitere während des Experiments gemessene Prozessparameter in die Modellierung aufgenommen. Zudem ist die Normalverteilungsannahme bezüglich der Schichteigenschaften sowie die lineare Struktur nicht notwendigerweise gegeben. Daher werden generalisierte lineare Modelle verwendet (Rehage et al., 2012a; Rudak et al., 2012).

Im folgenden wird zunächst der thermische Spritzprozess mit seinen Anwendungsgebieten kurz erläutert. Dabei wird auf die Einstellgrößen an der Maschine sowie Zielgrößen (Partikel- und Schichteigenschaften) eingegangen. Anschließend werden einige Herausforderungen bei der Modellierung des thermischen Spritzprozesses genannt und bisherige Ergebnisse zusammengefasst.

2.1. Thermische Spritzverfahren

Thermische Spritzprozesse werden angewendet, um eine Schicht auf einer Oberfläche aufzutragen, etwa um Verschleiß oder Korrosion vorzubeugen. Das Prinzip des thermischen Spritzprozesses ist standardisiert in der DIN EN 657. Spritzmaterial wird einem Spritzapparat, der als Hitzequelle dient, von außen oder innen in Form eines Pulvers

oder Drahtes zugeführt. Hier wird der Spritzzusatz (ganz oder teilweise) geschmolzen oder erhitzt, bis er weich wird. Anschließend beschleunigt ein Gasstrahl die heißen Partikel auf das zu beschichtende Substrat, wo die Partikel in Form von Flecken (sogenannte Splats) auftreffen, abkühlen und sich zu einer Schicht formen. Diese Schicht ist in der Regel zwischen 0.04 bis 3 mm dick. Abbildung 2.1 illustriert den Aufbau einer Schicht. Es finden sich hier auch ungeschmolzene und oxidierte Partikel.

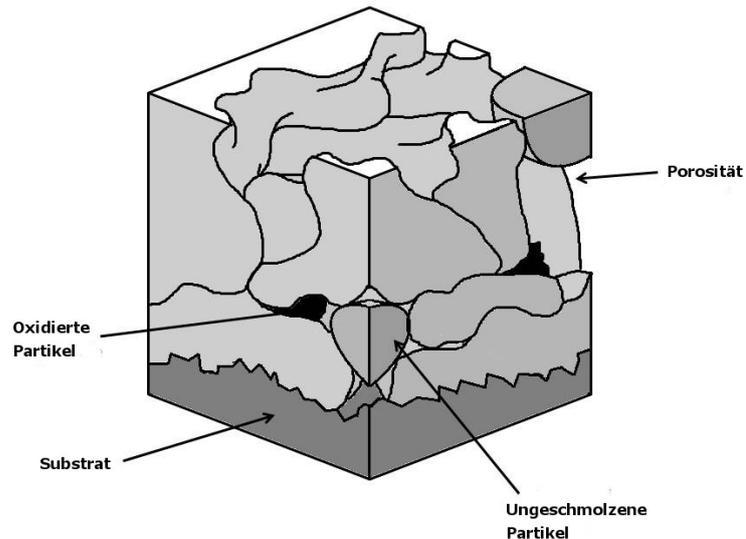


Abbildung 2.1: Aufbau einer Schicht

Die unterschiedlichen thermischen Spritzprozesse können nach der Hitzequelle klassifiziert werden. Die Üblichen sind dabei Flamspritzen, Hochgeschwindigkeitsflamspritzen (HVOF), Lichtbogenspritzen und Plasmaspritzen. Die dieser Arbeit zugrundeliegenden Versuche basieren auf einem Hochgeschwindigkeitsflammprozess, engl. High Velocity Oxygen Fuel Spray process (HVOF-Spritzprozess), siehe Abbildung 2.2, und wurden am Lehrstuhl für Werkstofftechnologie an der Technischen Universität Dortmund durchgeführt. Der HVOF-Spritzprozess wurde Anfang der 1980er Jahre entwickelt und gilt als einer der jüngsten thermischen Spritzprozesse. Mithilfe dieser Methode werden im Vergleich zu anderen Methoden deutlich höhere Partikelgeschwindigkeiten erreicht. Infolge-

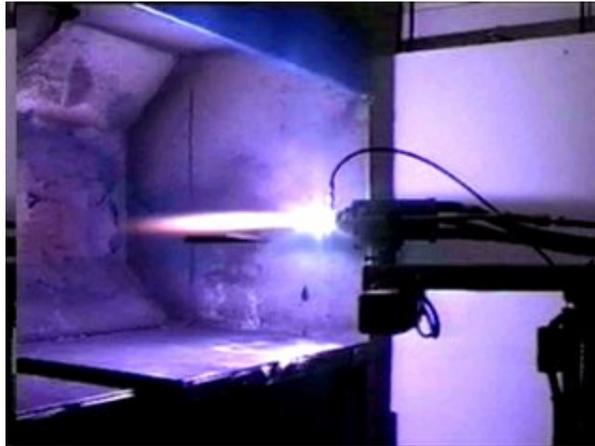


Abbildung 2.2: HVOF-Spritzprozess

dessen zeichnen sich die resultierenden Schichten durch geringere Porosität und höhere Verbundfestigkeit aus (Bach et al., 2006) und sind somit interessant für verschiedenste Anwendungen. Der Einsatz von HVOF-Spritzprozessen ist besonders für den Verschleiß- und Korrosionsschutz geeignet. Dabei stehen Karbidschichten aus hochverschleißfestem Wolframkarbid und Chromkarbidschichten im Vordergrund. Zudem gibt es viele Anwendungen, die metallische Schichtwerkstoffe verbinden.



Abbildung 2.3: Beschichtung einer Papierwalze

Industrielle Anwendungen können unterschiedlichsten Bereichen des allgemeinen Maschinenbaus zugeordnet werden. Für den Flüssigkeitserosions- und Kavitationsschutz finden HVOF-Schichten Anwendung bei beispielsweise Brennerdüsen, Sieben in der Salz-

industrie, Schaufelflächen von Wasserturbinen und Großventilatoren oder Papierwalzen (siehe Abbildung 2.3). Weitere Anwendungsfelder sind unter anderem die Luftfahrtindustrie (Gleitflächen von Hydraulikzylindern), der Apparatebau in der Gasindustrie und Petrochemie (Rohre, Ventile) oder der Automobilbau (Bremscheiben).

2.2. Beschreibung des Versuchsaufbaus

Es liegen für diese Arbeit Versuche vor, die in Rehage et al. (2012a), Rudak et al. (2012) und Borowski et al. (2014) beschrieben werden. Die Versuche wurden an einer Spritzpistole vom Typ *Wokajet 400 HVOF* durchgeführt (siehe Abbildung 2.4). Diese



Abbildung 2.4: Versuchsaufbau am Lehrstuhl für Werkstofftechnologie der TU Dortmund

Spritzpistole nutzt flüssiges Kerosin und reinen Sauerstoff für den Verbrennungsprozess. Ein Wolfram-Carbid-Cobald-Pulver (WC-12Co) vom Typ WOKA 3102 von Sulzer Metco wird für die Beschichtung eingesetzt. Die chemische Zusammensetzung des Pulvers ist in Tabelle 2.1 dargestellt. Die von Sulzer Metco ermittelte Korngröße des Pulvers liegt zwischen 15 und 45 μm . Dies wird bezeichnet mit $-45 + 15 \mu\text{m}$. Im Rahmen von Vorversuchen basierend auf Screening-Experimenten wurden die wesentlichsten Anlageparameter Kerosinfluss, Lambda (Verhältnis von Kerosin- und Sauerstofffluss), Spritz-

W	C	Co	Fe
82.3	5.48	12.13	0.09

Tabelle 2.1: Chemische Zusammensetzung des verwendeten Spritzpulvers in %

abstand und Geschwindigkeit der Förderscheibe ermittelt. Die Anlageparameter sowie interessierende Partikeleigenschaften und Schichteigenschaften sind in der Tabelle 2.2 zusammengefasst. Die Partikeleigenschaften werden mit einem Accuraspray g3 Analysesystem der Firma Tecnar erfasst. Die Größe und Form der Partikel wird von dem optischen Lasersystem VisiSizer N60 von Oxford Lasers gemessen. Es hat sich jedoch herausgestellt, dass die Anlageparameter bei diesem Pulver keinen messbaren Einfluss auf die Partikelgröße und Partikelform haben. Daher werden diese beiden Größen im Folgenden nicht mehr betrachtet.

Einstellgrößen X	Partikeleigenschaften Y	Schichteigenschaften Z
Kerosin (K)	Temperatur (Te)	Porosität(Po)
Lambda (L)	Geschwindigkeit (Ve)	Härte (Ha)
Spritzabstand (SOD)	Flambbreite (Wi)	Dicke (Th)
Geschwindigkeit der Förderscheibe (FDV)	Flammintensität (In)	Auftragswirkungsgrad(Dr)

Tabelle 2.2: Interessierende Anlageparameter sowie Partikel- und Schichteigenschaften im HVOF-Spritzprozess

2.3. Statistische Herausforderungen und bisherige Ergebnisse

In diesem Kapitel wird auf die statistischen Herausforderungen im Rahmen der Modellierung thermischer Spritzprozesse eingegangen. Dabei steht insbesondere fehlende Reproduzierbarkeit bedingt durch Tageseffekte (Temperaturschwankungen, Verschleiß)

im Vordergrund. Bisherige Ansätze zur Modellierung des thermischen Spritzprozesses beschränken sich überwiegend auf lineare Modelle, die Schichteigenschaften in Beziehung zu Anlageparametern setzen. Hier können keine tagesbedingten Störungen aufgenommen werden. Generalisierte lineare Modelle (McCullagh und Nelder, 1999) finden bisher keine Verwendung in der Literatur (Rudak et al., 2012).

Aufgrund von situationsbedingten, latenten Störgrößen ist die Herstellung einer reproduzierbaren, hohen Schichtqualität allein durch traditionelle Offline-Optimierung nahezu unmöglich. Aufgrund dessen werden Partikeleigenschaften im Flug in die Modellierung einbezogen, die unmittelbar gemessen werden können und Informationen über tagesbedingte Störungen enthalten. Der thermische Spritzprozess wird zweistufig modelliert über

$$X \rightarrow Y \rightarrow Z.$$

Dabei sollen auch mögliche multivariate Zusammenhänge zwischen Anlageparametern, Partikel- und Schichteigenschaften einfließen.

Eine weitere große Herausforderung stellt der geringe Datenumfang dar. Die Durchführung der Spritzversuche ist sehr zeitaufwändig und kostspielig. An einem Tag können maximal 30 Experimente durchgeführt werden. Schichteigenschaften können dabei lediglich durch aufwändige zerstörende Verfahren analysiert werden und dies kann mehrere Wochen dauern.

Zusammenfassung der wesentlichen Ergebnisse

Im folgenden werden die im Rahmen dieser Arbeit bereits entstandenen Veröffentlichungen zusammengefasst und eine Beschreibung der wesentlichen Ansätze und Ergebnisse angegeben.

a) M. Erdbrügge, S. Kuhnt, und N. Rudak. Joint optimization of independent multiple responses. *Quality and Reliability Engineering International*, 27:689–704, 2011

Zusammenfassung: In dieser Arbeit wird die Joint-Optimization-Methode (JOP-Methode) (eingeführt von Kuhnt und Erdbrügge (2004)) als Ansatz zur multikriteriellen Optimierung eines Produktionsprozesses mit einstellbaren Größen und mehreren unkorrelierten Zielgrößen durch Minimierung einer Risikofunktion für eine Reihe von Kostenmatrizen vorgestellt. Die Kostenmatrizen sind a priori nicht bekannt und es wird eine implementierbare Wahl von einer Reihe von Kostenmatrizen vorgeschlagen. Die Zielgrößen können nicht gleichzeitig optimiert werden und es kann höchstens ein „guter“ Kompromiss in Form eines Pareto-optimalen Punktes berechnet werden. Es konnte gezeigt werden, dass die JOP-Methode Pareto-optimale Punkte liefert. Die Ergebnisse werden in einem Joint-Optimization-Plot (JOP) zusammengefasst. Diese Methode wird in Kapitel 3 vorgestellt und im weiteren Verlauf in Kapitel 6 und 7 erweitert.

b) S. Kuhnt und N. Rudak. Simultaneous Optimization of Multiple Responses with the R Package JOP. *Journal of Statistical Software*, 54(9), 2013b

Zusammenfassung: In dieser Arbeit wird das R-Paket JOP vorgestellt, welches die JOP-Methode implementiert. Das R-Paket JOP bietet nach Eingabe der Datensätze eine Modellierung mit Selektion der interessierenden Zielgrößen sowie anschließende Minimierung der Risikofunktion für eine Reihe von Kostenmatrizen. Anschließend werden die Ergebnisse im Joint-Optimization-Plot zusammengefasst und der Anwender kann mithilfe der Maus einen geeigneten Kompromiss auswählen und erhält die entsprechenden Einstellgrößen. Weitere Details zur Implementierung der JOP-Methode folgen im Laufe der Arbeit in Kapitel 6.

c) 1. A. Rehage, N. Rudak, B. Hussong, S. Kuhnt, und W. Tillmann. Prediction of in-flight particle properties in thermal spraying with additive day-effects. SFB 823 discussion paper 06/12, TU Dortmund University, 2012a

2. A. Rehage, N. Rudak, B. Hussong, S. Kuhnt, und W. Tillmann. Einführung eines Tageseffekt-Schätzers zur Verbesserung der Vorhersage von Partikeleigenschaften in einem HVOF-Spritzstrahl. *Thermal Spray Bulletin*, 5(2), 2012b

Zusammenfassung: Um den zweistufigen Spritzprozess vollständig abzubilden, sind zunächst Modelle zwischen Prozessparametern und Partikeleigenschaften gesucht. In dieser Arbeit werden univariate generalisierte lineare Modelle für die Partikeleigenschaften aufgestellt und anhand des BIC-Kriteriums sowie Residualplots selektiert. Des Weiteren wird ein Ansatz zur Schätzung eines additiven Tageseffekts vorgestellt und erfolgreich auf einen realen Datensatz angewandt.

d) 1. N. Rudak, S. Kuhnt, B. Hussong, und W. Tillmann. On different strategies for the prediction of coating properties in a HVOF process. SFB 823 discussion paper 29/12, TU Dortmund University, 2012

2. B. Hussong, W. Tillmann, S. Kuhnt, N. Rudak, und A. Rehage. Vergleich unterschiedlicher Strategien zur Prozessmodellierung beim HVOF thermischen Spritzen. In *Tagungsband zum 15. Werkstofftechnischen Kolloquium*, pages 68–73, Chemnitz, 2012

Zusammenfassung: Als Weiterführung von Rehage et al. (2012a) werden verschiedene Strategien zur Vorhersage von Schichteigenschaften untersucht. Es werden vier unterschiedliche Methoden zur Vorhersage von Schichteigenschaften vorgestellt und miteinander verglichen. Diese Ansätze unterscheiden sich grundlegend in der Art der Eingangsdaten, die im Modell enthalten sind.

Beim klassischen direkten Modell werden die Schichteigenschaften anhand der Anlageparameter vorhergesagt. Das indirekte Modell nutzt für die Vorhersage der Schichteigenschaften die Partikeleigenschaften. Das zusammengesetzte Modell bildet zunächst eine Vorhersage \hat{Y} der Partikeleigenschaften (Modelle aus Rehage et al. (2012a)) und anschließend eine Vorhersage der Schichteigenschaften \hat{Z} auf Basis der geschätzten Partikeleigenschaften \hat{Y} . Weiterhin nutzt das hybride Modell für die Vorhersage der Schichteigenschaften sowohl die Anlageparameter als auch die Partikeleigenschaften.

Zunächst werden die Modelle ausgewählt und anhand von PRESS-Werten und Residualplots verglichen. Anschließend werden die Vorhersagegenauigkeiten der Modelle anhand von vier Versuchen an einem anderen Tag untersucht. Dabei sticht keines der Modelle

als klarer Favorit heraus. Zu bemerken ist jedoch, dass das zusammengesetzte Modell sehr gut abschneidet und in jedem Fall das direkte Modell von den anderen Ansätzen übertroffen wird.

e) 1. N. Rudak, S. Kuhnt, und E. Riccomagno. Numerical algebraic fan of a design for statistical model building. SFB 823 discussion paper 04/13, TU Dortmund University, 2013

2. N. Rudak, S. Kuhnt, und E. Riccomagno. Numerical algebraic fan of a design for statistical model building. *Statistica Sinica*, 2015b. doi: 10.5705/ss.2014.264

Zusammenfassung: In dem zweistufigen Prozess $X \rightarrow Y \rightarrow Z$ wird ein Standardversuchsplan auf der Ebene der Anlageparameter festgelegt. Sollen nun Modelle zwischen Y und Z gebildet werden, so liegt kein Standardplan mehr auf der Partikelebene vor. Daher ist nicht klar, welche Modelle identifizierbar sind. Einen möglichen Weg identifizierbare Modelle zu bestimmen, bietet die algebraische Statistik. Erstmals wird ein numerischer Buchberger-Möller-Algorithmus eingesetzt, um identifizierbare Modelle für den thermischen Spritzprozess zu bestimmen.

f) N. Rudak, S. Kuhnt, und E. Riccomagno. Modeling of a Thermal Spraying Process by Gaussian Chain Graphs. *Quality Technology and Quantitative Management*, 11(1): 85–98, 2014

Zusammenfassung: In dieser Arbeit werden vier unterschiedliche Methoden zur Anpassung eines sogenannten Chain Graph Model (CG) vorgestellt. Zu jeder dieser Methoden steht ein frei erhältliches Programmpaket zur Verfügung. Nach einer Präsentation der grundlegenden Theorie folgt eine Einführung in die entsprechenden Programmpakete. Anschließend werden die Methoden auf einen realen Datensatz resultierend aus thermischen Spritzexperimenten angewandt. Die unterschiedlichen Methoden liefern unterschiedliche Ergebnisse, auch wenn sich Gemeinsamkeiten finden lassen. In diesem Artikel wird ein erster Eindruck von multivariaten Zusammenhängen in dem zweistufigen thermischen Spritzprozess deutlich.

g) M. Borowski, N. Rudak, B. Hussong, D. Wied, S. Kuhnt, und W. Tillmann. On- and offline detection of structural breaks in thermal spraying processes. *Journal of Applied Statistics*, 41(5):1073–1090, 2014

Zusammenfassung: HVOF-Spritzsysteme enthalten bereits Alarmsysteme, die bei Störungen anschlagen. Gewisse Störungen, wie etwa dem Ausfall eines Pulverförderers oder Druckverlust des Fördergases, werden nicht sofort erkannt und können unter Umständen einen Einfluss auf die Schichtqualität haben. In dieser Arbeit werden vier unterschiedliche Störungen bewusst herbeigeführt und die Partikeldaten vor und nach der Störung aufgezeichnet. Die Störungen machen sich in den Zeitreihen der Partikeleigenschaften bemerkbar, sowohl in der Veränderung des absoluten Werts (Sprung oder Trendveränderung) als auch in der Varianz. Daher werden ein neu entwickeltes Verfahren zur Sprung- und Trenderkennung (SCARM-Verfahren) und der neu entwickelte Fluktuationstest zur Erkennung von Veränderungen in der Varianz eingesetzt. Das Ergebnis des Fluktuationstests kann durch Sprünge in der Zeitreihe verfälscht werden. Dem kann mit der vorangehenden Anwendung eines Filters zur Beseitigung des Sprungs vorgebeugt werden. Hier ist das Ergebnis auch wieder abhängig von der gewählten Fensterbreite. Daher wird in einer Simulationsstudie der empirische Fehler erster Art des Fluktuationstests in Abhängigkeit der Sprunggröße und der gewählten Fensterbreite untersucht. Dazu wird standardnormalverteiltes Rauschen erzeugt und in der Mitte der Zeitreihe wird ein Sprung von absoluter Größe 0.5, 1, 2 addiert. Anschließend wird ein Filter mit Fensterbreiten 11, 31 und 51 angewandt und der Fluktuationstest ausgeführt. Ohne Filtern liegt der empirische Fehler erster Art bereits bei einem Sprung von 0.5 bei 0.18. Mit Filtern liegt der empirische Fehler erster Art unabhängig von der Fensterbreite unter 0.05.

h) N. Rudak, B. Hussong, und S. Kuhnt. Simultaneous optimization of multiple correlated responses with application to a thermal spraying process. *Quality and Reliability Engineering International*, 2015a. akzeptiert zur Veröffentlichung

Zusammenfassung: In dieser Arbeit wird die JOP-Methode auf korrelierte Zielgrößen

erweitert. Dies geschieht zum einen durch Verwendung multivariater Modelle zur Schätzung des Erwartungswertvektors und der Kovarianzmatrix eines Zufallsvektors $Y|x$ in Abhängigkeit eines Kovariablenvektors x . Es werden verschiedene Modelle vorgestellt und in einer Simulationsstudie verglichen. Zum anderen werden nichtdiagonale Kostenmatrizen betrachtet und eine konkrete Wahl dieser Matrizen eingeführt.

Im weiteren Verlauf der Arbeit wird auf die Prozessoptimierung mithilfe der JOP-Methode weiter eingegangen. Dazu werden zunächst der bedingte Erwartungswertvektor $E(Y|x)$ und die bedingte Kovarianzmatrix $Cov(Y|x)$ in Abhängigkeit von Kovariablen modelliert. Erdbrügge et al. (2011) und Kuhnt und Rudak (2013b) modellieren die einzelnen Zielgrößen univariat und gehen von einer diagonalen Kovarianzmatrix aus. In dieser Arbeit wird die Methode hinsichtlich der Aufnahme einer multivariaten Modellierung sowohl von $E(Y|x)$ als auch von $Cov(Y|x)$ im nichtdiagonalen Fall erweitert.

3. Simultane Optimierung multipler Zielgrößen

Gerade in vielen technischen Anwendungen besteht das Ziel darin, mehrere Zielgrößen gleichzeitig zu optimieren. Diese Zielgrößen können beispielsweise von Maschineneinstellungen abhängig sein. Dabei kann die Aufgabe darin bestehen, einzelne Zielgrößen zu minimieren oder zu maximieren oder Einstellungen zu finden, die im Erwartungswert bestimmte Zielwerte erreichen. Zudem besteht ein weiteres Kriterium darin, eine Lösung zu finden, die auch gleichzeitig die Varianz minimiert. Sind die Zielgrößen möglicherweise untereinander korreliert, so sollte auch diese Möglichkeit in der Modellierung oder Optimierung aufgenommen werden, um eine möglichst realitätsnahe Lösung zu erhalten. Diese Problemstellung läuft auf ein multikriterielles Optimierungsproblem hinaus und oft können nicht alle Zielgrößen gleichzeitig optimiert werden. Es kann lediglich ein "guter" Kompromiss gewählt werden.

Eine einfache und aufschlussreiche Methode der Darstellung der zu optimierenden Zielfunktionen stellt die Verwendung von Konturplots dar (Myers et al., 2009), die die Niveaulinien der aus Daten geschätzten Zielfunktionen darstellen. Ein beispielhafter Konturplot ist in Abbildung 3.1 dargestellt. Auf der horizontalen und vertikalen Achse sind

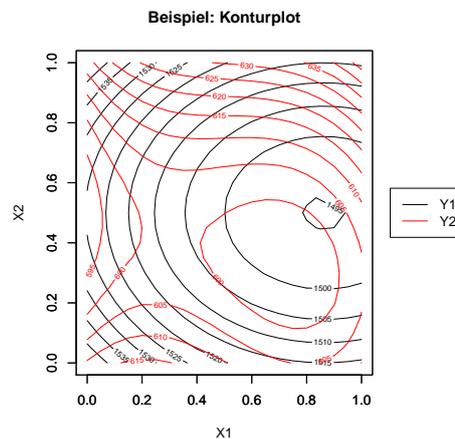


Abbildung 3.1: Beispielhafter Konturplot

hier die Designvariablen x_1 und x_2 abgetragen. Die Niveaulinien der geschätzten Erwartungswerte

tungswertfunktionen der Zielgröße Y_1 sind schwarz und die von Y_2 sind rot dargestellt. Sollen hier beide Zielgrößen minimiert werden, so ist zu erkennen, dass der minimale Wert für Y_1 (1495) auf der rechten Seite mittig erwartet wird und für Y_2 (595) auf der linken Seite mittig. Ein möglicher Kompromiss ist auf der rechten Seite mittig in Betracht zu ziehen, da für Y_2 auch hier ein niedriger Wert (600) erreicht wird. Bei mehr als drei Designvariablen $x \in \mathbb{R}^k$ müssen bei der Darstellung durch Konturplots $k - 2$ Designvariablen konstant gehalten werden. Daher ist hier diese Analyse schwierig und es müssen geeignetere Optimierungsmethoden herangezogen werden.

In dieser Arbeit interessieren dabei vor allem Methoden, die neben der Optimierung der Erwartungswerte mehrerer Zielgrößen versuchen auch deren Varianz zu minimieren, sowie Korrelationen zu beachten. Dieser Anforderung werden die meisten Methoden bisher nur teilweise gerecht.

Der Ansatz von Chiao und Hamada (2001) basiert auf der Annahme einer multivariaten Normalverteilung und zielt darauf ab eine Einstellung zu finden, die die Wahrscheinlichkeit dafür, dass die Zielvariablen innerhalb einer festgelegten Region liegen, maximiert. Aufgrund der Normalverteilungsannahme sind die logarithmierten empirischen Varianzen sowie die transformierten empirischen Korrelationen approximativ normalverteilt mit konstanter Varianz (Hamada und Balakrishnan, 1998). Daher werden die einzelnen Erwartungswerte, sowie die transformierten empirischen Varianzen und Korrelationen mithilfe gewöhnlicher linearer Modelle abgebildet.

Viele weitere Ansätze zur simultanen Optimierung multipler Zielgrößen basieren auf Wünschbarkeiten oder Verlustfunktionen. Ein umfangreicher Überblick über verschiedenste Methoden zur Lösung von multikriteriellen Optimierungsproblemen ist in Ardakani und Wulff (2013) zu finden. Eine spezielle Zusammenstellung und ein Vergleich wünschbarkeitsbasierter Methoden ist in Costa et al. (2011) dargestellt. Verlustbasierte Methoden beruhen meist auf der Bestimmung einer optimalen Lösung für eine bestimmte Kostenmatrix. Die hier weiterentwickelte JOP-Methode (Kuhnt und Erdbrügge, 2004) betrachtet eine ganze Reihe von Kostenmatrizen und ermöglicht dem Anwender un-

terschiedliche Lösungen zu finden, die sowohl die Varianz der Zielgrößen minimieren als auch die Zielwerte möglichst gut erreichen sollen. Der Anwender kann anschließend einen geeigneten Kompromiss anhand eines bestimmten Kriteriums oder grafisch mithilfe eines sogenannten Joint Optimization Plots wählen. Im weiteren Verlauf der Arbeit werden einige Eigenschaften dieser Methode nachgewiesen. Des Weiteren wird die JOP-Methode in dieser Arbeit auf multiple korrelierte Zielgrößen erweitert.

Im folgenden bezeichne $Y = (Y_1, \dots, Y_p)^T$ einen p -dimensionalen Zielgrößenvektor, der beispielsweise die Merkmale eines Produkts repräsentiert. Des Weiteren sei $\tau = (\tau_1, \dots, \tau_p)^T$ der Vektor der angestrebten Zielwerte und $x = (x_1, \dots, x_k)^T$ seien die kontrollierbaren Einflussgrößen wie etwa Maschineneinstellungen.

3.1. Wünschbarkeitsbasierte Methoden

Wünschbarkeitsbasierte Methoden (Derringer und Suich, 1980) sind sehr beliebt bei Anwendern und werden überwiegend in der Industrie zur Optimierung multipler Zielgrößen eingesetzt (Ardakani und Wulff, 2013). Die Idee besteht darin, jede Zielgröße Y_i in eine individuelle Wünschbarkeit d_i zu transformieren, die nur Werte zwischen 0 und 1 annimmt. Die Zielgröße hat den Zielwert erreicht für $d_i = 1$ und ist inakzeptabel für $d_i = 0$ (Myers et al., 2009). Anschließend werden die individuellen Wünschbarkeiten d_i in eine globale Wünschbarkeit transformiert, z.B.

$$D = (d_1 \cdot \dots \cdot d_p)^{\frac{1}{p}} \text{ oder } D = (d_1^{w_1} \cdot \dots \cdot d_p^{w_p})^{\frac{1}{\sum w_i}}.$$

Die globale Wünschbarkeit D ist bei derartigen geometrischen Produkten bereits 0, wenn eine einzelne Wünschbarkeit $d_i = 0$ ist. Einen Überblick über weitere wünschbarkeitsbasierte Ansätze ist in Costa et al. (2011) zu finden.

In Wu (2005) wird eine wünschbarkeitsbasierte Methode vorgestellt, die auch Korrelationen zwischen den Zielgrößen berücksichtigt. Es bezeichne $\Sigma(x) = Cov(Y|x)$ die bedingte Kovarianzmatrix und $\mu(x) = E(Y|x)$ den bedingten Erwartungswert. Wu (2005) kon-

struiert Wünschbarkeiten wie folgt für $i = 1, \dots, p$

$$d_i(x) = \exp(-l_i(x)) \quad (3.1)$$

mit $l_i(x) = ([C]_{ii}[\Sigma(x)]_{ii} + ([\mu(x)]_i - [\tau]_i)^2)$ und für $i \neq j \in \{1, \dots, p\}$

$$d_{ij}(x) = \begin{cases} 1 & , \text{ falls gilt } l_{ij}(x) < 0, \\ \exp(-l_{ij}) & , \text{ sonst.} \end{cases} \quad (3.2)$$

mit $l_{ij}(x) = [C]_{ij}([\Sigma(x)]_{ij} + ([\mu(x)]_i - [\tau]_i)([\mu(x)]_j - [\tau]_j))$. Die Wünschbarkeit $d_i(x)$ bezieht sich hier auf die Zielgröße Y_i und $d_{ij}(x)$ bezieht sich simultan auf die Zielgrößen Y_i und Y_j . Die zu erreichenden Zielwerte werden mit $\tau \in \mathbb{R}^p$ bezeichnet und $C \in \mathbb{R}^{p \times p}$ stellt eine Kostenmatrix dar. Maximiert wird nun

$$D(x) = \left\{ \prod_{i=1}^p \prod_{i < j}^p d_i(x) d_{ij}(x) \right\}^{\frac{1}{p}}.$$

Die Maximierung von $D(x)$ führt zu Einstellungen, die die Wünschbarkeitsfunktionen $d_i(x)$ und $d_{ij}(x)$ gleichzeitig maximieren. Falls für $x^* = \operatorname{argmax}_x D(x)$ der Wert $D(x^*) = 1$ erreicht wird, so gilt $l_i(x^*) = 0$ und $l_{ij}(x^*) = 0$ für alle $i, j = 1, \dots, p$.

Wu (2009) stellt einen robusten dynamischen wünschbarkeitsbasierten Ansatz zur Optimierung multipler Zielgrößen durch Einbeziehung von Störvariablen, sogenannten Signalvariablen und Korrelationen vor. Die Optimierung geschieht auf Grundlage von doppelt-exponentiellen Wünschbarkeitsfunktionen, die im Folgenden zunächst kurz erläutert werden. Soll ein Zielwert T erreicht werden, so lautet die doppelt-exponentielle Wünschbarkeit

$$d(x) = \begin{cases} \exp(-c_1 |\mu(x) - T|^{\alpha_1}), & -\infty < \mu(x) \leq T \\ \exp(-c_2 |\mu(x) - T|^{\alpha_2}), & T \leq \mu(x) < \infty \end{cases}.$$

Die doppelt-exponentielle Wünschbarkeitsfunktion ist insbesondere geeignet, wenn obere oder untere Grenzen pro Zielgröße nicht sinnvoll gewählt werden können. Diese Wünschbarkeitsfunktion nimmt stets positive Werte $d(x) > 0$ an. Die Wahl der Werte $c_1, c_2, \alpha_1, \alpha_2$

bestimmt die Form der Wünschbarkeitsfunktion. Soll die Zielgröße möglichst minimale Werte annehmen, so lautet die Wünschbarkeit $d(x) = \exp(-c|\mu(x) - L|^\alpha)$ für den kleinstmöglichen Wert $L \leq \mu(x) < \infty$, der erreicht werden kann. Bei Maximierung wird die Wünschbarkeitsfunktion

$$d(x) = \begin{cases} 1 - \frac{\exp(-c\mu(x)^\alpha)}{\exp(-cL^\alpha)} & L \leq \mu(x) < \infty \\ 0 & \mu(x) < L \end{cases}$$

gewählt. Neben den kontrollierbaren Einflussgrößen x seien Störvariablen z_l , $l = 1, \dots, r$, gegeben. Zudem werden sogenannte Signalvariablen M_1, \dots, M_r betrachtet und es wird angenommen, dass der Zielwertvektor funktional von diesen Signalvariablen abhängt. Die beobachtete Zielgröße $\mu_{ikl}(x)$ ist der Wert der beobachteten i -ten Zielvariable $Y_i|x$ bei Einstellung x , Störgröße z_l und Signalvariable M_k . Der Vektor $\tau(M_k)$ bezeichne also den Zielwertvektor bezüglich Signalvariable M_k , $k = 1, \dots, r$. Korrelationen zwischen den Zielgrößen $Y_i|x$ und $Y_j|x$ werden durch die **korrelierte Wünschbarkeit**

$$d_{ij}(x) = \min \left\{ 1, \left(\prod_{k=1}^n \prod_{l=1}^r \exp(-c_{ij}(\mu_{ikl}(x) - [\tau(M_k)]_i)(\mu_{jkl}(x) - [\tau(M_k)]_j)) \right)^{\frac{1}{nr}} \right\}$$

berücksichtigt. Zusammen mit

$$d_i(x) = \left\{ \prod_{k=1}^n \prod_{l=1}^r \exp(-c_i|\mu_{ikl}(x) - [\tau(M_k)]_i|^{\alpha_i}) \right\}^{\frac{1}{n}}$$

wird

$$D(x) = \left\{ \prod_{i=1}^p \prod_{i < j}^p d_i(x)d_{ij}(x) \right\}^{\frac{1}{p}}$$

maximiert.

Weitere Methoden basieren auf einer speziellen Gewichtung der Wünschbarkeitsindizes (Wonggasem, 2012). Zudem wird teilweise eine Hauptkomponentenanalyse als Vorstufe eingesetzt (Salmasnia et al., 2012, 2013b,a). Dabei werden die Zielgrößen zunächst normalisiert und anschließend eine Hauptkomponentenanalyse durchgeführt. Die Zielwerte werden ebenso normalisiert und in die Hauptkomponenten übergeführt. Nur die

ausgewählten Hauptkomponenten fließen in den weiteren Optimierungsprozess ein. Bei diesem Vorgehen werden Korrelationen berücksichtigt und die Anzahl der Zielgrößen unter Umständen reduziert, falls einige der Hauptkomponenten vernachlässigt werden können, ohne dass wesentliche Informationen verloren gehen. In Salmasnia und Bashiri (2015) werden Korrelationen zwischen den Wünschbarkeiten $d_i(x)$ und $d_j(x)$ bezüglich den Zielgrößen Y_i und Y_j zunächst geschätzt und bei der weiteren Optimierung berücksichtigt. Diese Ansätze benötigen in der Regel Wiederholungen pro Versuchseinstellung.

3.2. Verlustbasierte Methoden

Verlustbasierte Methoden beruhen auf der Betrachtung eines Verlustes, der im einfachsten Fall durch die Abweichung einer univariaten Zielgröße Y von ihrem Zielwert τ entsteht, etwa $L(Y) = c(Y - \tau)^2$, wobei $c > 0$ die durch Abweichung entstehenden Kosten beschreibt (Taguchi et al., 1989; Chandra, 2001). Diese Verlustfunktion kann über eine Taylor-Entwicklung auf $L(Y)$ mit

$$L(Y) = L(\tau + Y - \tau) = L(\tau) + \frac{(Y - \tau)}{1!} \frac{\partial}{\partial Y} L(Y) + \frac{(Y - \tau)^2}{2!} \frac{\partial^2}{\partial^2 Y} L(Y) + \dots$$

hergeleitet werden. Die Verlustfunktion soll in τ Null ergeben und ebenda ihr Minimum annehmen. Durch diese beiden Annahmen an die Verlustfunktion fallen die ersten beiden Terme weg. Nach Vernachlässigung der höheren Terme ab Grad 3 erhält man mit $c := \frac{\partial^2 L(Y)}{\partial^2 Y}$ die Verlustfunktion $L(Y) = c(Y - \tau)^2$.

In Antony (2000) wird eine Methode zur Optimierung multipler Zielgrößen vorgestellt, die auf den Ideen von Taguchi basiert und diese mit der Hauptkomponentenanalyse verbindet. Es werden für $Y|x$ Verluste der Art $L_{ij} = c([Y|x_j]_i - \tau_i)$ betrachtet, wobei τ_i dem zu erreichenden Zielwert bezogen auf die i -te Zielgröße entspricht. Anschließend werden die Verluste normalisiert durch $\tilde{L}_{ij} = \frac{L_{ij} - L_{i,*}}{L_{i,*} - L_{i,*}}$ mit $L_{i,*} = \min_j L_{ij}$ und $L_{i,*} = \max_j L_{ij}$. Für die normalisierten Verlustwerte \tilde{L}_{ij} wird eine Hauptkomponentenanalyse durchgeführt. Anschließend werden Einstellungen bestimmt, die die wichtigste Hauptkomponente maximieren.

Mukhopadhyay und Khuri (2008) kombinieren multivariate GLM mit einer multivariaten Verlustfunktion. Für einen Kovariablenvektor $x \in \mathbb{R}^k$ wird angenommen, dass die Verteilung des bedingten Zielvariablenvektors $Y|x$ zu einer q -parametrischen Exponentialfamilie gehört (siehe Kapitel 4.1). Aufgrund dieser Annahme ist die Kovarianzmatrix $\text{Cov}(Y|x)$ nicht notwendigerweise diagonal. Hier fließen bereits mögliche Korrelationen in die Modellierung ein. Mit einer geeigneten invertierbaren Linkfunktion $g: \mathbb{R}^q \rightarrow \mathbb{R}^q$ wird ein linearer Prädiktor $Z(x)^T\beta$ mit dem Erwartungswert $\mu(x) = E(Y|x)$ über $g(\mu(x)) = Z(x)^T\beta$ in Beziehung gesetzt. Dabei ist $Z(x) \in \mathbb{R}^{p \times q}$ eine geeignete Matrix und $\beta \in \mathbb{R}^q$ der zu schätzende Koeffizientenvektor, die geschätzte Erwartungswertfunktion ist damit $\hat{\mu}(x) = g^{-1}(Z^T(x)\hat{\beta})$.

Minimiert wird nun eine Distanzfunktion der Art

$$\rho(\hat{\mu}(x), \tau(\hat{\beta})) = \left[(\hat{\mu}(x) - \tau(\hat{\beta}))^T \widehat{\text{Cov}}(\hat{\mu}(x))^{-1} (\hat{\mu}(x) - \tau(\hat{\beta})) \right]^{\frac{1}{2}}. \quad (3.3)$$

Die Distanzfunktion (3.3) berechnet den quadratischen Abstand der geschätzten Erwartungswertfunktion $\hat{\mu}(x)$ vom Zielwertvektor $\tau(\hat{\beta})$ gewichtet durch die Inverse der geschätzten Kovarianzmatrix $\widehat{\text{Cov}}(\hat{\mu}(x))$. Mithilfe einer Taylor-Entwicklung wird eine approximative Schätzung der Kovarianzmatrix durch

$$\widehat{\text{Cov}}(\hat{\mu}(x)) = \frac{\widehat{\mu(x)}}{\eta(x)^T} \left(Z(x)^T \widehat{\text{Cov}}(\hat{\beta}) Z(x) \right) \frac{\widehat{\mu(x)^T}}{\eta(x)}$$

erhalten. Der Zielwertvektor $\tau(\hat{\beta})$ enthält die individuell erreichbaren Optimalwerte $\tau_j(\hat{\beta})$ für jede geschätzte Erwartungswertfunktion $\hat{\mu}_j(x) = g^{-1}(Z^T(x)\hat{\beta})$. Die Schätzung $\hat{\beta}$ von β hängt von den Zufallsvektoren Y_i ab und damit sind $\tau_j(\hat{\beta})$ selbst auch Zufallsvariablen. Die Variabilität von $\tau(\hat{\beta})$ soll im Folgenden bei der Optimierung berücksichtigt werden. Bezeichne nun $\tau(\beta)$ den Vektor der unbekanntenen wahren Erwartungswertfunktion $\mu(x)$. Die Distanzfunktion

$$\rho(\hat{\mu}(x), \tau(\beta)) = \left[(\hat{\mu}(x) - \tau(\beta))^T \widehat{\text{Var}}(\hat{\mu}(x))^{-1} (\hat{\mu}(x) - \tau(\beta)) \right]^{\frac{1}{2}} \quad (3.4)$$

kann folglich nicht direkt minimiert werden, sondern lediglich eine obere Schranke, die wie folgt aussieht

$$\min_x \rho(\hat{\mu}(x), \tau(\beta)) \leq \min_x \max_{\tau \in D_\tau} \rho(\hat{\mu}(x), \tau), \quad (3.5)$$

wobei D_τ eine Konfidenzregion für $\tau(\beta)$ zum Niveau α bezeichnet. Die rechte Seite der Ungleichung (3.5) wird nun minimiert.

Andere Methoden beruhen auf der Minimierung einer erwarteten Verlustfunktion (Pignatiello, 1993; Vining, 1998; Ko et al., 2005). Pignatiello (1993) betrachtet die multivariate Verlustfunktion

$$\text{loss}(Y) = (Y - \tau)^T C (Y - \tau) \quad (3.6)$$

mit einer Kostenmatrix $C \in \mathbb{R}^{p \times p}$ und einem Vektor $\tau \in \mathbb{R}^p$ der zu erreichenden Zielwerte. Die Einträge der Kostenmatrix können durch bekannte Verluste festgelegt werden, siehe dazu Pignatiello (1993). Die Elemente außerhalb der Diagonalen geben den Verlust an, wenn gleichzeitig zwei verschiedene Zielvariablen die Zielwerte verfehlen. Die Diagonalwerte können als Gewichte für die jeweilige Zielvariable interpretiert werden. Minimiert werden soll die geschätzte erwartete Verlustfunktion

$$\begin{aligned} \hat{R}(x) &= \hat{E}(\text{loss}(Y)) = \hat{E} \left((Y - \tau)^T C (Y - \tau) \right) \\ &= \text{Spur}(C \hat{\Sigma}(x)) + (\hat{\mu}(x) - \tau)^T C (\hat{\mu}(x) - \tau) \end{aligned} \quad (3.7)$$

mit $\hat{\Sigma}(x) = \widehat{\text{Cov}}(Y|x)$ und $\hat{\mu}(x) = \hat{E}(Y|x)$. Pignatiello (1993) verwendet für die Schätzung von $E(Y|x)$ und $\text{Cov}(Y|x)$ das arithmetische Mittel und die empirische Kovarianzmatrix. Zur Schätzung der Risikofunktion muss zudem eine Kostenmatrix C vorab fest gewählt werden. Ko et al. (2005) erweitern den Ansatz um die Unsicherheit sowohl der Zielgröße als auch des Modells. Vining (1998) betrachtet die erwartete Verlustfunktion für \hat{Y} wie folgt

$$\hat{R}_{\hat{Y}}(x) = \hat{E}(\text{loss}(\hat{Y})) = \text{Spur}(C \widehat{\Sigma}_{\hat{Y}}(x)) + (\hat{\mu}(x) - \tau)^T C (\hat{\mu}(x) - \tau),$$

wobei sich hier $\widehat{\Sigma}_{\hat{Y}}(x)$ auf die Kovarianzmatrix der Vorhersage $\hat{\mu}(x)$ bezieht, während in Pignatiello (1993) die Kovarianzmatrix der Zielgröße Y betrachtet wird. In allen genannten Methoden wird die Kostenmatrix einmal festgelegt und anschließend die Optimierung durchgeführt. Die gesamte Kovarianzmatrix ist Teil der Risikofunktion und somit werden bei Verwendung einer nichtdiagonalen Kostenmatrix auch Korrelationen

berücksichtigt, sofern die gesamte Kovarianzmatrix $\Sigma(x)$ geschätzt wird. Dagegen zielt die JOP-Methode darauf ab, die geschätzte Risikofunktion (3.7) zu schätzen und anschließend für eine ganze Folge von Kostenmatrizen C zu minimieren.

3.3. Verlustbasierte JOP Methode

In Kuhnt und Erdbrügge (2004) wird eine verlustbasierte robuste Methode zur Optimierung multipler Zielgrößen eingeführt, die auf die Optimierung der Risikofunktion (3.7) abzielt. In einem sogenannten Joint Optimization Plot werden die Ergebnisse der Optimierung für Sequenzen von Kostenmatrizen grafisch zusammenfasst und dem Anwender ermöglicht, einen Kompromiss zu finden.

Die JOP-Methode basiert auf der Minimierung der geschätzten Risikofunktion (3.7)

$$\hat{R}(x) = \text{Spur}(C\hat{\Sigma}(x)) + (\hat{\mu}(x) - \tau)^T C (\hat{\mu}(x) - \tau)$$

für eine ganze Folge von Kostenmatrizen C_1, \dots, C_m , die die Gewichtung der einzelnen Zielgrößen zueinander festlegen. Eine simultane Modellierung der unbekanntenen Erwartungswerte und der Varianzen kann durch individuelle doppelt generalisierte lineare Modelle erfolgen (Smyth, 1989), die in Kapitel 4.2 vorgestellt werden. Bei Verwendung von diagonalen Kostenmatrizen werden keine Kovarianzmodelle benötigt. Die mittels geeigneter Minimierungsmethoden gefundenen Lösungen von

$$\min_{x \in X} \hat{R}(x)$$

für eine Sequenz von Kostenmatrizen C_1, \dots, C_m können mithilfe des sogenannten Joint Optimization Plots dargestellt werden, um einen geeigneten Kompromiss auszuwählen.

Eine exemplarische Darstellung des Joint Optimization Plot (JOP) ist in Abbildung 3.2 dargestellt. Der JOP besteht aus zwei Grafiken. Die x-Achse bezieht sich in beiden Grafiken auf eine Folge von Kostenmatrizen C_1, \dots, C_m . Diese seien durch einen reellen so genannten Streckungsvektor parametrisiert. Die Einträge des Streckungsvektors sind

auf der x-Achse abgetragen und jeder einzelne Wert kann einer Kostenmatrix zugeordnet werden. Die y-Achse der rechten Grafik bezieht sich auf die Werte der Zielvariablen Y_r , $r = 1, \dots, p$. In dieser Grafik finden sich für jede Kostenmatrix C_i , $i = 1, \dots, m$, die Pareto optimale Lösung, interpoliert durch die durchgezogene Linie. Das Band um die durchgezogenen Linien ist ebenfalls interpoliert und stellt eine Unsicherheit von der Größenordnung \pm geschätzte Standardabweichung von Y_r dar. Die horizontalen gestrichelten Linien stehen für die zu erreichenden Zielwerte. Die linke Grafik stellt die zugehörigen Prozessparameterwerte (auf der y-Achse) gegen die Einträge des Streckungsvektors (auf der x-Achse) dar.

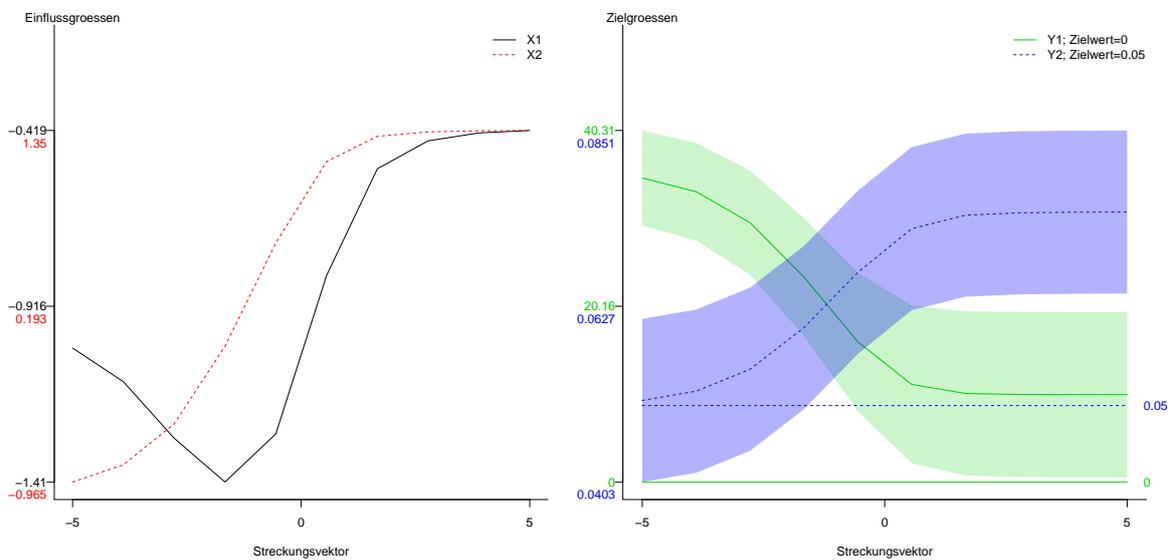


Abbildung 3.2: Beispielhafter Joint Optimization Plot

Der Anwender soll alle Ergebnisse auf einen Blick sehen können und mit dem nötigen Hintergrundwissen über die zugrundeliegende Anwendung in der Lage sein, einen geeigneten Kompromiss mit den entsprechenden Einstellungen x zu finden. Ein guter Kompromiss wäre folglich da zu finden, wo die Zielgrößen nahe an ihren Zielwerten liegen und gleichzeitig das Unsicherheitsband möglichst klein ist. Abbildung 3.2 verdeutlicht die typische Situation eines multikriteriellen Optimierungsproblems, wo die Zielgrößen jeweils gegenläufig sind. Auf der linken Seite der rechten Grafik liegt die blaue Zielgröße nahe an ihrem Zielwert, wobei die grüne Zielgröße weit von ihrem Zielwert entfernt ist.

Auf der rechten Seite der Grafik ist es umgekehrt.

Im weiteren Verlauf der Arbeit wird in Kapitel 6 und 7 eine konkrete Wahl der Kostenmatrizen vorgestellt. Des Weiteren werden einige Eigenschaften der JOP-Methode nachgewiesen und die Implementierung in Form des R-Pakets JOP (Kuhnt und Rudak, 2013b) präsentiert. Abschließend erfolgt die Erweiterung der JOP-Methode auf multiple korrelierte Zielgrößen. Dazu werden jedoch vorab Schätzer $\hat{\mu}(x)$ für den Erwartungswertvektor $E(Y|x)$ und $\hat{\Sigma}(x)$ für die Kovarianzmatrix $Cov(Y|x)$ benötigt, um die multivariate Struktur der Zielgröße einzubeziehen.

4. Simultane Modellierung des Erwartungswerts und der Varianz

In diesem Kapitel werden zunächst generalisierte lineare Modelle, die den Erwartungswert $E(Y|x)$ modellieren, vorgestellt (McCullagh und Nelder, 1999). Anschließend werden doppelt generalisierte lineare Modelle (Smyth, 1989) eingeführt, die eine gleichzeitige Schätzung des Erwartungswerts $E(Y|x)$ und der Varianz der Zielgröße $Var(Y|x)$ in Abhängigkeit von Kovariablen erlauben. Diese Modelle finden in der JOP-Methode (Kuhnt und Erdbrügge, 2004) Verwendung mit dem Ziel Parametereinstellungen zu finden, für die der Erwartungswert der Zielgröße gewünschte Werte mit möglichst kleiner Varianz erreicht. Es wird angenommen, dass die Verteilung der Zielgröße einer Exponentialfamilie entstammt.

4.1. Generalisierte lineare Modelle

Nelder und Wedderburn (1972) führen generalisierte lineare Modelle (GLM) ein und vereinen damit eine ganze Reihe von Modellen, die als GLM dargestellt werden können und bis dahin unabhängig voneinander behandelt wurden. Ein Spezialfall des GLM ist das klassische lineare Modell. Für eine Zufallsvariable Y und einen Kovariablenvektor $x \in \mathbb{R}^k$ geht das klassische lineare Modell

$$Y = f(x)^T \beta + \epsilon$$

von einem normalverteilten Fehler $\epsilon \sim N(0, \sigma^2)$ aus. Dabei ist $\beta \in \mathbb{R}^q$ der zu schätzende Parametervektor und es ergibt sich $Y \sim N(f(x)^T \beta, \sigma^2)$. Weiterhin wird $f(x) \in \mathbb{R}^q$ als Designvektor bezeichnet. Beispielsweise kann der Designvektor für $x = (x_1, x_2)^T$ durch $f(x) = (x_1, x_1^2, x_2)^T$ gegeben sein.

Damit lässt sich das lineare Modell wie folgt beschreiben.

1. *Zufällige Komponente:* Die Zufallsvariable $Y|x$ ist normalverteilt mit Erwartungs-

wert $\mu(x) = E(Y|x)$ und Varianz σ^2 .

2. *Strukturkomponente*: Der lineare Prädiktor ist gegeben durch $\eta = f(x)^T \beta$.
3. *Link*: Der lineare Prädiktor ist über die Beziehung $\mu = \eta$ mit dem Erwartungswert $E(Y|x)$ verknüpft.

Der Übergang zum GLM besteht in der Verallgemeinerung der zufälligen Komponente und des Links. Die dritte Komponente wird erweitert durch

$$\eta = g(\mu)$$

mit einer beliebigen monotonen und differenzierbaren Funktion g . Im Fall des klassischen linearen Modells entspricht $g(\cdot)$ der Identität. Des Weiteren darf die Verteilung von $Y|x$ einer Exponentialfamilie entstammen (Witting, 1985).

Definition 4.1 (Exponentialfamilie) Sei Ω ein Ereignisraum, \mathcal{A} eine σ -Algebra auf Ω und \mathcal{T} ein Parameterraum. Sei $P_\theta : \mathcal{A} \rightarrow [0, 1]$ für $\theta \in \mathcal{T}$ ein Wahrscheinlichkeitsmaß. Dann heißt $\mathcal{P} = \{P_\theta, \theta \in \mathcal{T}\}$ k -parametrische Exponentialfamilie, wenn ein dominierendes und σ -endliches Maß α , eine Zahl $k \in \mathbb{N}$ und reellwertige \mathcal{A} -messbare Funktionen T_1, \dots, T_k, c auf Ω existieren mit

$$\frac{dP_\theta}{d\alpha}(Y) = a(\theta) e^{\sum_{i=1}^k [\theta]_i T_i(Y)} c(Y) = a(\theta) e^{\theta^T T(Y)} c(Y)$$

für $\theta \in \mathcal{T}$ und $T(Y) = (T_1(Y), \dots, T_k(Y))$.

Damit $\frac{dP_\theta}{d\alpha}(Y)$ eine Dichte darstellt, ist $a(\theta)$ ein normalisierender Faktor von der Form

$$a(\theta) = \frac{1}{\int e^{\sum_{i=1}^k [\theta]_i T_i(Y)} c(Y) d\alpha}.$$

Der Vektor θ wird *natürlicher* Parameter genannt. Für $T(Y) = Y$ in Definition 4.1 spricht man von einer *einfachen* Exponentialfamilie, deren Elemente eine ν -Dichte $d\nu = \frac{\partial P_\theta}{\partial \alpha}$ bezüglich des σ -endlichen dominierenden Maßes α der Form

$$\frac{dP_\theta}{d\alpha}(Y) = a(\theta) e^{\theta^T Y} c(Y) = e^{\theta^T Y - b(\theta)} c(Y) \quad (4.1)$$

mit $b(\theta) = \log a(\theta)$ haben.

Durch die Einführung eines weiteren sogenannten **Dispersionsparameters** ψ (Fahrmeir und Tutz, 2001) wird (4.1) zu

$$ED(Y; \theta, \psi) := e^{\frac{\theta^T Y - b(\theta)}{\psi}} c(Y, \psi). \quad (4.2)$$

Es gelten die folgenden Beziehungen (Fahrmeir und Tutz, 2001)

$$E_{\theta}(Y) = \frac{\partial b(\theta)}{\partial \theta} \quad (4.3)$$

$$Cov_{\theta}(Y) = \psi \frac{\partial^2 b(\theta)}{\partial \theta \partial \theta^T} := \psi V(\mu) \quad (4.4)$$

mit einer sogenannten *Kovarianzfunktion* $V(\mu)$, die von dem Erwartungswert μ abhängt. Einfache Exponentialfamilien werden vollständig durch ihre Varianzfunktion beschrieben (Jorgensen, 1987, Theorem 1).

Eine Reihe von häufig auftretenden Verteilungsklassen, wie etwa die Normalverteilung oder die Gamma-Verteilung, sind in einparametrischen Exponentialfamilien enthalten. Ein generalisiertes lineares Modell für eine univariate Zielgröße Y wird nun wie folgt definiert (Fahrmeir und Tutz, 2001; McCullagh und Nelder, 1999).

Definition 4.2 (Generalisiertes Lineares Modell (GLM)) *Sei die Zufallsvariable Y die Zielgröße und $X \in \mathbb{R}^k$ der Kovariablenvektor. Ein GLM wird durch folgende Komponenten beschrieben.*

- *Zufällige Komponente: Der Zufallsvektor $Y|x$ besitzt eine bedingte Verteilung bei gegebenem $X = x$, die einer einfachen einparametrischen Exponentialfamilie entstammt und deren Dichte $ED(Y; \theta, \psi)$ sich wie in (4.2) darstellen lässt bezüglich eines σ -endlichen dominierenden Maßes.*
- *Strukturkomponente: Der lineare Prädiktor ist gegeben durch $\eta = f(x)^T \beta$.*
- *Link: Der lineare Prädiktor ist über die Beziehung $g(\mu) = \eta$ mit dem Erwartungswert verknüpft, wobei $g(\cdot)$ eine monotone differenzierbare Funktion ist.*

Für $\theta = \eta$ spricht man von einer *natürlichen* Linkfunktion. Multivariate Erweiterungen

des GLM für multiple beispielsweise multinomialverteilte Zielgrößen werden in Fahrmeir und Tutz (2001) vorgestellt.

Seien Stichprobenvariablen $Y_1|x_1, \dots, Y_n|x_n$ mit zugehörigen Kovariablenvektoren $x_i \in \mathbb{R}^k$ und Designvektoren $f(x_i) \in \mathbb{R}^q$ gegeben, wobei die Verteilung von $Y_i|x_i$ einer einparametrischen natürlichen Exponentialfamilie entstamme. Sei $X \in \mathbb{R}^{n \times q}$ die Designmatrix mit Zeilen $f(x_i)$ für $i = 1, \dots, n$. Zudem überführe die Abbildung ϕ den natürlichen Parameter in den Erwartungswertvektor, $\phi : \theta \rightarrow \mu = \frac{\partial b(\theta)}{\partial \theta}$. Mit $\mu_i = g(\eta_i)$ und $\eta_i = f(x_i)^T \beta$ (siehe Definition 4.2) lautet der i-te Anteil der log-Likelihoodfunktion $l = \sum_{i=1}^n l_i(\beta)$ in β

$$l_i(\beta) = \frac{Y_i \phi^{-1}(g^{-1}(f(x_i)^T \beta)) - b(\phi^{-1}(g^{-1}(f(x_i)^T \beta)))}{\psi} + \ln(c(Y_i, \psi)).$$

Mit der Varianzfunktion $V(\cdot)$ ergibt sich der Gradient der log-Likelihoodfunktion mithilfe der Kettenregel zu (McCullagh und Nelder, 1999)

$$\frac{\partial l}{\partial \beta} = \sum_{i=1}^n \frac{Y_i - \mu_i}{\psi V(\mu_i)} \frac{\partial g^{-1}}{\partial \eta} \Big|_{f(x_i)^T \beta} f(x_i). \quad (4.5)$$

Der Maximum-Likelihood-Schätzer für β kann mithilfe einer gewichteten Kleinste-Quadrate-Methode (IWLS) bestimmt werden (McCullagh und Nelder, 1999).

Sei $\hat{\beta}$ der Maximum-Likelihood-Schätzer für den Stichprobenumfang n und β der wahre Parametervektor. Unter gewissen Regularitätsbedingungen (Fahrmeir und Kaufmann, 1985; Fahrmeir und Tutz, 2001) ist $\hat{\beta}$ konsistent für β und asymptotisch normalverteilt mit $n^{\frac{1}{2}}(\hat{\beta} - \beta) \stackrel{asympt.}{\sim} N(0, I(\beta)^{-1})$. Hier ist $I(\beta) = -E\left(\frac{\partial^2 l}{\partial \beta \partial \beta^T}\right)$ die Fischer'sche Informationsmatrix. Nelder und Wedderburn (1972) zeigen, dass der IWLS-Algorithmus mit der *Fischer-Scoring-Methode* übereinstimmt, die zu der Iterationsvorschrift

$$\beta^{(t+1)} = \beta^{(t)} + I(\beta^{(t)})^{-1} \frac{\partial l(\beta)}{\partial \beta} \Big|_{\beta=\beta^{(t)}}$$

führt. Für einen Startvektor $\beta^{(0)}$ (beispielsweise der gewöhnliche Kleinste-Quadrate-Schätzer) wird diese Iterationsvorschrift bis zum Abbruch fortgeführt.

Zur Modellwahl kann die Devianz oder bekannte Kriterien wie das AIC-Kriterium oder das BIC-Kriterium herangezogen werden. Gewöhnliche Residuen können nicht benutzt

werden, da die Varianz von dem Erwartungswert abhängt und somit entsprechend standardisiert werden muss. Daher können beispielsweise sogenannte Pearson-Residuen $r_p = \frac{Y - \mu}{\sqrt{V(\mu)}}$ angewandt werden. Für weitere Details zur Modellwahl- und Diagnose sei auf Fahrmeir und Tutz (2001) oder McCullagh und Nelder (1999) verwiesen.

Ist der Dispersionsparameter bekannt, so kann $\text{Var}(Y)$ direkt über die Varianzfunktion geschätzt werden. Ist der Dispersionsparameter ψ unbekannt, so kann er mithilfe der ML-Methode, der Momentenmethode oder über (Hinkley et al., 1991)

$$\hat{\psi} = \frac{1}{n - r} \sum_{i=1}^n \frac{(Y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)},$$

geschätzt werden, wobei r dem Rang der Designmatrix X entspricht. Soll der Dispersionsparameter auch von den Regressoren abhängen können, so ist eine Erweiterung des Modells nötig.

4.2. Doppelt generalisierte lineare Modelle

Das Vorliegen eines möglicherweise nicht konstanten Dispersionsparameters, wie etwa durch den Einfluss von Kovariablen auf den Dispersionsparameter, wird durch das klassische GLM nicht berücksichtigt. Die Varianz $\text{Var}(Y) = \phi V(\mu)$ ist bereits durch die Verteilungsannahme festgelegt, siehe Gleichung (4.4).

Doppelt generalisierte lineare Modelle (DGLM) erlauben die gleichzeitige Schätzung des Erwartungswerts und der Varianz in Abhängigkeit von Kovariablen und werden in Smyth (1989) eingeführt. Für Stichprobenvariablen $Y_1|x_1, \dots, Y_n|x_n$, deren Verteilung einer einfachen einparametrischen Exponentialfamilie entstammt, wird die Dichte (4.2) umgeschrieben zu

$$a(Y, \psi) e^{-\frac{1}{2\psi} d(Y, \mu(\theta))}, \quad (4.6)$$

wobei $d(Y, \mu(\theta))$ ein Maß für den Abstand zwischen Y und $\mu(\theta)$ ist (Smyth und Verbyla, 1999). Die Erweiterung von Smyth (1989) besteht nun in der Annahme ungleicher

Dispersionsparameter ψ_i für $i = 1, \dots, n$. Die log-Likelihoodfunktion basierend auf (4.6) lautet nun

$$l(\theta, \psi_1, \dots, \psi_n) = \sum_{i=1}^n -\frac{1}{2\psi_i} d(Y_i, \mu_i) + \tilde{a}(Y_i, \psi_i) \quad (4.7)$$

mit $\tilde{a}(y, \psi) := \log(a(y, \psi))$.

Schätzung mit der Maximum-Likelihood-Methode (ML)

Für die Maximum-Likelihood-Schätzung der Parameter im DGLM seien im Folgenden $f^e(x_1), \dots, f^e(x_n) \in \mathbb{R}^q$ sowie $f^d(x_1), \dots, f^d(x_n) \in \mathbb{R}^w$ zugehörige Designvektoren. Sei $X \in \mathbb{R}^{n \times q}$ die Designmatrix mit Zeilen $f^e(x_i)$, $i = 1, \dots, n$ und $Z \in \mathbb{R}^{n \times w}$ die Designmatrix mit Zeilen $f^d(x_i)$, $i = 1, \dots, n$. Nun werde der Erwartungswert $E(Y_i) = \mu_i$ über

$$\mu_i = g^{-1}(f^e(x_i)^T \beta)$$

und der Erwartungswert $E(D_i) = \delta_i$ für Zufallsvariablen $D_i = d(Y_i, \mu_i)$ über

$$\delta_i = h^{-1}(f^d(x_i)^T \gamma)$$

mit geeigneten Linkfunktionen $g(\cdot)$ und $h(\cdot)$ modelliert. Für feste ψ_i ist die Gleichung (4.7) eine log-Likelihoodfunktion für ein GLM für den Erwartungswert μ_i , wobei ψ_i^{-1} die Gewichte sind und der Dispersionsparameter 1 beträgt. Dieses wird als *Erwartungswertsubmodell* bezeichnet. Anwendung der Fischer-Scoring-Methode ergibt die Iterationsvorschrift

$$\beta^{(k+1)} = (X^T W_1 X)^{-1} X^T W_1 r^{(1)} \quad (4.8)$$

mit $W_1 = \text{diag}\left(\left(\frac{\partial g}{\partial \mu_i} \Big|_{\mu_i}\right)^{-2} V(\mu_i)^{-1}\right)$, $r_i^{(1)} = \frac{\partial g}{\partial \mu_i} \Big|_{\mu_i} (Y_i - \mu_i) + g(\mu_i)$ und $\mu_i = g^{-1}(f^e(x_i)^T \beta^{(k)})$.

Für die Dispersion ergibt sich nach Gleichung (4.7) entsprechend das *Dispersionssubmodell* für ψ_i bei festen μ_i als GLM mit Realisation d_i von Zufallsvariablen $D_i = d(Y_i, \mu_i)$,

Erwartungswerten $E(D_i) = \delta_i$, Varianzen $Var(D_i) = \psi^d \tilde{V}(\delta_i) = 2\tilde{V}(\delta_i)$ und Dispersionsparameter $\psi^d = 2$ (Smyth, 1989). Damit folgt die Iterationsvorschrift

$$\gamma^{(k+1)} = (Z^T W_2 Z)^{-1} Z^T W_2 r^{(2)} \quad (4.9)$$

mit $W_2 = \text{diag}\left(2\left(\frac{\partial h}{\partial \delta_i} |_{\delta_i}\right)^2 \tilde{V}(\delta_i)\right)^{-1}$, $r_i^{(2)} = \frac{\partial h}{\partial \delta_i} |_{\delta_i} (d_i - \delta_i) + h(\delta_i)$ und $\delta_i = h^{-1}(f^d(x_i)^T \gamma^{(k)})$.

Die Schätzung erfolgt nun durch abwechselnde Lösung der Gleichungen (4.8) und (4.9). Für gegebenes $\hat{\gamma}^{(k)}$ wird ein $\hat{\beta}^{(k+1)}$ über die Vorschrift (4.8) und für gegebenes $\hat{\beta}^{(k+1)}$ ein $\hat{\gamma}^{(k+1)}$ über die Vorschrift (4.9) bestimmt.

Unter Annahme einer Normalverteilung und inverser Normalverteilung für die Variablen Y_i folgen die Größen D_i einer \mathcal{X}_1^2 -Verteilung und damit entspricht das Dispersionsmodell einer Gammaverteilung (Smyth, 1989). Bei Annahme einer Gamma-Verteilung haben die Devianzkomponenten die Gestalt $D_i = 2 \{\log(\Gamma(\psi_i^{-1})) - \Psi(\psi_i^{-1})\}$, wobei Ψ der Digammafunktion entspricht. Die Erwartungswerte $E(D_i)$ und Varianzen $Var(D_i)$ werden mithilfe der Digammafunktion berechnet (Smyth, 1989). Des Weiteren kann die Dichte (4.2) mithilfe der sogenannten Sattelpunktapproximation $c(y, \psi) \approx \{2\pi\psi V(\mu)\}^{-\frac{1}{2}}$ (Smyth und Verbyla, 1999) angenähert werden. Damit kann gezeigt werden, dass die Devianzen D_i approximativ $\psi_i \mathcal{X}_1^2$ -verteilt sind. Die Sattelpunktapproximation ist exakt für die Normalverteilung und die inverse Normalverteilung.

Sind die Normalverteilung, inverse Normalverteilung oder die Gammaverteilung als Annahme nicht gerechtfertigt, so führt die Festlegung einer Varianzfunktion $V(\mu) = \mu^p$ zu sogenannten Tweedy-Modellen als Spezialfälle der Verteilungsfunktionen mit Dichte (4.2) (Smyth und Dunn, 2008). Für $p \notin (0, 1)$ gehört zu jeder solchen Varianzfunktion eine Verteilungsfunktion mit Dichte (4.2) (Jorgensen, 1987; Smyth und Dunn, 2008). Die zugehörige Dichte kann über Fourierreihen (Smyth und Dunn, 2008) oder über die Sattelpunktapproximation (Smyth und Verbyla, 1999) bestimmt werden. Für die Bestimmung von Tweedy-Dichten liegt das R-Paket `tweedie` vor (Dunn, 2014). Eine restriktierte Maximum-Likelihood-Schätzung für DGLM führt zu (approximativ) unverzerrten Schätzern der Varianz (Smyth und Verbyla, 1999).

5. Simultane Modellierung des Erwartungswertvektors und der Kovarianzmatrix

In praktischen Situationen werden oft mehrere Zielgrößen in einem Experiment gleichzeitig gemessen. Ausgehend von den Modellen in Kapitel 4 können die Zielgrößen einzeln durch GLM oder DGLM modelliert werden. Mit den geschätzten Modellen können dann mithilfe von Wünschbarkeiten oder der JOP-Methode (Kuhnt und Rudak, 2013b) Parametereinstellungen gefunden werden, die gewisse Zielwerte mit möglichst kleiner Varianz erreichen. Bei diesem Vorgehen bleiben jedoch mögliche Korrelationen zwischen den Zielgrößen unberücksichtigt.

In diesem Kapitel werden daher Modelle betrachtet und erweitert, die sowohl den multivariaten Erwartungswert als auch die gesamte Kovarianzmatrix in Abhängigkeit von Kovariablen modellieren. Zunächst ist es wichtig positive Definitheit der Kovarianzmatrix zu sichern. Positive Definitheit kann durch zusätzliche Restriktionen erreicht werden, was jedoch schwierig umzusetzen ist, wie in Pinheiro und Bates (1996) herausgestellt wird. Eine andere Möglichkeit besteht in einer Parametrisierung der Kovarianzmatrix, die positive (Semi-)Definitheit und bestenfalls eine sinnvolle Interpretierbarkeit der zu schätzenden Parameter zulässt. Im Anschluss kann eine unrestringierte Schätzung der Kovarianzmatrix durch beispielsweise die Maximierung einer Likelihoodfunktion erfolgen. Einen Überblick über mögliche unrestringierte Parametrisierungen wird in Pinheiro und Bates (1996) und Pourahmadi (2011) gegeben.

In Pourahmadi (1999, 2000) wird ein multivariates generalisiertes lineares Modell sowohl für den bedingten Erwartungswert $E(Y|x)$ als auch die Kovarianzmatrix $\Sigma(x) = \text{Cov}(Y|x)$ entwickelt. Das Modell für die Kovarianzmatrix basiert auf der Cholesky-Zerlegung und hängt somit von der Ordnung der Einträge des Zielvariablenvektors Y ab. Die Schätzung erfolgt durch Anwendung der Maximum-Likelihood-Methode. Des Weiteren wird in Huang et al. (2006) eine penalisierte Maximum-Likelihood-Schätzung sowie in Papageorgiou (2012) eine restringierte Maximum-Likelihood-Schätzung vorgestellt. Der

Ansatz von Pourahmadi (1999) ist insbesondere sinnvoll bei longitudinalen Daten aufgrund der vorgegebenen Ordnung. Da in der Anwendung auf thermische Spritzprozesse keine Ordnung vorliegt, wird diese Methode im Folgenden nicht weiter berücksichtigt.

Chiao und Hamada (2001) bilden separate gewöhnliche lineare Modelle für die Erwartungswerte, Varianzen und Korrelationen durch Verwendung von arithmetischen Mitteln, empirischen Varianzen und empirischen Korrelationen als Zielgrößen. Dafür werden jedoch für jede Versuchseinstellung Wiederholungen benötigt und daher wird diese Art der Modellierung im folgenden außer Acht gelassen.

In diesem Kapitel werden drei unterschiedliche Methoden vorgestellt und erweitert (Chiu et al., 1996; Hoff und Niu, 2012; Yin et al., 2010), die eine simultane Schätzung des Erwartungswertvektors und der Kovarianzmatrix erlauben. Die Erweiterung des Modells von Yin et al. (2010) erfolgt durch die Verwendung einer multivariaten Kernfunktion. Die Erweiterung des Modells von Chiu et al. (1996) besteht in der Verwendung einer allgemeinen Linkfunktion für den Erwartungswertvektor. Des Weiteren wird das Modell von Hoff und Niu (2012) mit einer alternativen Parametrisierung dargestellt und durch den Einsatz einer allgemeinen Linkfunktion für den Erwartungswertvektor erweitert. Mithilfe der alternativen Parametrisierung wird der Weg zur Modellselektion sowohl für das Erwartungswertmodell als auch für das Kovarianzmatrixmodell geebnet. Abschließend werden die Modelle in einer Simulationsstudie verglichen.

Im folgenden bezeichne erneut Y einen p -dimensionalen Zufallsvektor und $x \in \mathbb{R}^k$ sei wie in Kapitel 4 ein Kovariablenvektor. Der Zufallsvektor $Y|x$ folge stets einer multivariaten Normalverteilung mit Erwartungswertvektor $\mu(x) = E(Y|x)$ und einer Kovarianzmatrix $\Sigma(x) = \text{Cov}(Y|x)$. Seien weiterhin Realisationen y_1, \dots, y_n von n Stichprobenvariablen Y_1, \dots, Y_n mit zugehörigen Kovariablenvektoren x_1, \dots, x_n gegeben.

5.1. Multivariater Nadaraya-Watson-Kernschätzer

Yin et al. (2010) entwickeln einen Nadaraya-Watson-Kernschätzer zur Schätzung der Kovarianzmatrix. Ausgehend von einem Zufallsvektor $Y \in \mathbb{R}^p$ und einer zugehörigen Zufallsvariable $X \in \mathbb{R}$ wird angenommen, dass $Y|X = x$ einer Verteilung mit Erwartungswert $\mu(x) = (\mu_1(x), \dots, \mu_p(x))^T$ und Kovarianzmatrix $\Sigma(x)$ mit Einträgen $[\Sigma(x)]_{ij} = \sigma_{ij}(x)$ für $i, j = 1, \dots, p$ folgt. Die Funktionen $\mu(x)$ und $\Sigma(x)$ werden als glatt angenommen. Sei $K(x)$ eine symmetrische Kernfunktion und $K_h(x) = h^{-1}K(\frac{x}{h})$ die skalierte Kernfunktion mit Bandbreite $h > 0$. Eine symmetrische Kernfunktion $K(x)$ muss folgende Bedingungen (Silverman, 1998) erfüllen

$$\int K(x) dx = 1, \quad \int xK(x) dx = 0, \quad \int x^2K(x) dx \neq 0. \quad (5.1)$$

Seien nun Stichprobenvariablen Y_1, \dots, Y_n mit Kovariablenvektoren x_1, \dots, x_n gegeben und sei x ein beliebiger Punkt des zulässigen Bereichs der Kovariable. Unter Annahme einer multivariaten Normalverteilung für $Y|x$ erhält man Schätzer durch Minimierung des Ausdrucks

$$\frac{1}{n} \sum_{i=1}^n \left[(Y_i - \mu(x))^T \Sigma^{-1}(x) (Y_i - \mu(x)) - \log(|\Sigma^{-1}(x)|) \right] K_h(x_i - x) \quad (5.2)$$

bezüglich $\mu(\cdot)$ und $\Sigma(\cdot)$ für festes x .

Der Ausdruck $\left[(Y_i - \mu(x))^T \Sigma^{-1}(x) (Y_i - \mu(x)) - \log(|\Sigma^{-1}(x)|) \right]$ entspricht der negativen logarithmierten Dichte einer multivariaten Normalverteilung, wobei die Konstante 2π vernachlässigt wurde. Damit kann (5.2) als logarithmierte Likelihoodfunktion angesehen werden, wobei die einzelnen Summanden durch die Kernfunktion $K_h(x_i - x)$ gewichtet werden. Durch Minimierung ergeben sich die folgenden Kernschätzer

$$\hat{\mu}(x) = \left(\sum_{i=1}^n K_h(x_i - x) Y_i \right) \left(\sum_{i=1}^n K_h(x_i - x) \right)^{-1} \quad \text{und} \quad (5.3)$$

$$\hat{\Sigma}(x) = \left[\sum_{i=1}^n K_h(x_i - x) (Y_i - \hat{\mu}(x_i)) (Y_i - \hat{\mu}(x_i))^T \right] \left(\sum_{i=1}^n K_h(x_i - x) \right)^{-1}. \quad (5.4)$$

Unter der Annahme einer ganzen Reihe von Regularitätsbedingungen zeigen Yin et al. (2010) asymptotische Normalität des Kernschätzers für die Kovarianz.

Der Ansatz von Yin et al. (2010) wird im Folgenden durch die Verwendung einer multivariaten Kernfunktion $K(x)$ mit $\int K(x) dx = 1$ mit $x \in \mathbb{R}^k$ für die Schätzer (5.3) und (5.4) erweitert. Betrachtet wird eine skalierte Kernfunktion $K_h(x) = h^{-k}K(\frac{1}{h}x)$ mit einer einzelnen Bandbreite $h > 0$ (Silverman, 1998). Somit wird jeder Eintrag von x gleich gewichtet. Mögliche Kernfunktionen sind beispielsweise die Dichte der multivariaten Normalverteilung mit Erwartungswertvektor $(0, \dots, 0)^T \in \mathbb{R}^k$ und der Einheitsmatrix als Kovarianzmatrix wie folgt

$$K_1(x) = (2\pi)^{-\frac{k}{2}} \exp\left(-\frac{1}{2}x^T x\right). \quad (5.5)$$

Weitere Möglichkeiten sind beispielsweise

$$K_2(x) = \begin{cases} 3\pi^{-1}(1 - x^T x)^2 & , \text{falls } x^T x < 1 \\ 0 & , \text{andernfalls} \end{cases}$$

$$K_3(x) = \begin{cases} \frac{1}{2}c_k^{-1}(k + 2)(1 - x^T x) & , \text{falls } x^T x < 1 \\ 0 & , \text{andernfalls} \end{cases}$$

$$K_4(x) = \begin{cases} 4\pi^{-1}(1 - x^T x)^3 & , \text{falls } x^T x < 1 \\ 0 & , \text{andernfalls.} \end{cases}$$

Hier bezeichnet c_k das Volumen der k -dimensionalen Sphäre $\{x \in \mathbb{R}^k \mid \|x\| \leq 1\}$. Im weiteren Verlauf der Arbeit wird stets angenommen, dass der Zufallsvektor multivariat normalverteilt ist und daher wird die Kernfunktion (5.5) verwendet, die allen Punkten x mit gleichem Abstand zu x_i identische Gewichte zuordnet. Alternative Wahlen können dennoch in Betracht gezogen werden, was jedoch in dieser Arbeit nicht erfolgt. Die Bandbreite $h > 0$ wird ausgewählt durch Minimierung von

$$CV_{\Sigma}(h) = \frac{1}{n} \sum_{i=1}^n \left[(Y_i - \hat{\mu}_{-i}(x_i))^T \hat{\Sigma}_{-i}(x_i)^{-1} (Y_i - \hat{\mu}_{-i}(x_i)) - \log(|\hat{\Sigma}_{-i}(x_i)^{-1}|) \right].$$

Hier ist $\hat{\Sigma}_{-i}(x_i)$ die Schätzung der Kovarianzmatrix $Y|x_i$ auf Basis von (5.4), wobei $\hat{\Sigma}(x)$ in Gleichung (5.4) auf Grundlage der beobachteten Daten mit Ausnahme der i -ten Beobachtung berechnet wurde. Analoges gilt für $\hat{\mu}_{-i}(x_i)$. Insgesamt liegt somit ein multivariater Nadaraya-Watson-Kernschätzer für den Erwartungswertvektor und die Kovarianzmatrix in Abhängigkeit von Kovariablen vor.

5.2. Schätzer basierend auf dem Matrix-Logarithmus-Modell mit einer allgemeinen Linkfunktion

Die Methode von Chiu et al. (1996) basiert auf der Matrix-Logarithmus-Parametrisierung (Bellman, 1970). Eine positiv definite Matrix Σ besitzt ausschließlich positive Eigenwerte $\lambda_i, i = 1, \dots, p$. Folglich existiert eine **Matrix-Logarithmus-Parametrisierung** der Art

$$\Sigma = \exp(A) := \sum_{i=0}^{\infty} \frac{A^i}{i!}. \quad (5.6)$$

mit einer reellen symmetrischen Matrix $A \in \mathbb{R}^{p \times p}$. Umgekehrt ist für eine symmetrische reelle Matrix $A \in \mathbb{R}^{p \times p}$ die Matrix $C = \exp(A)$ positiv definit (Chiu et al., 1996). Eine symmetrische reelle Matrix A besitzt eine **Spektralzerlegung** der Form

$$A = U\Delta U^T, \quad (5.7)$$

wobei $\Delta = \text{diag}(\lambda_1, \dots, \lambda_p)$ die Eigenwerte der Matrix A auf der Diagonalen enthält und die Spalten von U den Eigenvektoren der Matrix A entsprechen. Mit $A = U\Delta U^T$ und wegen der Orthogonalitätseigenschaft der Transformationsmatrizen U gilt $A^i = U\Delta^i U^T$ mit $[\Delta^i]_{jj} = \lambda_j^i$ für alle $i \in \mathbb{N}$ und $j = 1, \dots, p$. Damit ergibt sich $\Sigma = U \exp(\Delta) U^T$, wobei $\exp(\Delta)$ die Werte $\exp(\lambda_j)$ für $j = 1, \dots, p$ auf der Diagonalen enthält. Der Matrix-Logarithmus ist damit gegeben durch

$$\log(\Sigma) = U\Delta U^T$$

mit logarithmierten Eigenwerten der Matrix Σ auf der Diagonalen der Matrix $\log(\Delta)$ stehen. Die Abbildung $\log : \Sigma \rightarrow \log(\Sigma)$ ist bijektiv. Daher wird jeder Kovarianzmatrix

genau ein entsprechender Logarithmus zugeordnet. Der Vorteil dieser Parametrisierung besteht vor allem darin, dass die Einträge der Matrix A unrestringiert modelliert werden können. Eine wichtige Eigenschaft für die Matrix $\Sigma = \exp(A)$ mit einer reellen symmetrischen Matrix A , die für die spätere Schätzung gebraucht wird, ist mit den obigen Bezeichnungen

$$\begin{aligned} \log(|\Sigma|) &= \log(|U \exp(\Delta) U^T|) = \log(|\exp(\Delta)|) = \log\left(\prod_{i=1}^p \exp(\lambda_i)\right) \\ &= \sum_{i=1}^n \lambda_i = \text{Spur}(U \Delta U^T) = \text{Spur}(A). \end{aligned} \quad (5.8)$$

Im folgenden wird davon ausgegangen, dass die einzelnen Erwartungswerte $\mu_j(x)$ von Designvektoren $f_j(x) \in \mathbb{R}^{q_j}$ über die Beziehung

$$\mu_j(x) = g_j^{-1}(f_j(x)^T \beta_j)$$

mit $\beta_j \in \mathbb{R}^{q_j}$ ($j = 1, \dots, p$) abhängen. Hier ist $g_j^{-1} : \eta_j \in \mathbb{R} \rightarrow g_j^{-1}(\eta_j) \in \mathbb{R}$ mit $\eta_j = f_j(x)^T \beta_j$ eine geeignete Linkfunktion (siehe Kapitel 4.1) für $j = 1, \dots, p$. Setze nun

$$g^{-1} = \begin{pmatrix} g_1^{-1} \\ \vdots \\ g_p^{-1} \end{pmatrix} : \eta = \begin{pmatrix} \eta_1 \\ \vdots \\ \eta_p \end{pmatrix} \in \mathbb{R}^p \longrightarrow g^{-1}(\eta) = \begin{pmatrix} g_1^{-1}(\eta_1) \\ \vdots \\ g_p^{-1}(\eta_p) \end{pmatrix} \in \mathbb{R}^p$$

und

$$Z_1(x) = \begin{pmatrix} f_1(x) & 0_{q_1} & 0_{q_1} & \dots & 0_{q_1} & 0_{q_1} \\ 0_{q_2} & f_2(x) & 0_{q_2} & \dots & 0_{q_2} & 0_{q_2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0_{q_{p-1}} & 0_{q_{p-1}} & 0_{q_{p-1}} & \dots & f_{p-1}(x) & 0_{q_{p-1}} \\ 0_{q_p} & 0_{q_p} & 0_{q_p} & \dots & 0_{q_p} & f_p(x) \end{pmatrix} \in \mathbb{R}^{q \times p} \quad (5.9)$$

mit $q = q_1 + \dots + q_p$.

Das Erwartungswertmodell in Chiu et al. (1996) wird als linear angenommen mit $\mu(x) = Z_1(x)^T \beta$. Als Erweiterung des Modells von Chiu et al. (1996) wird hier eine Linkfunktion

zur Modellierung des Erwartungswertvektors integriert. Der Erwartungswertvektor wird über

$$\mu(x) = g^{-1}(Z_1(x)^T \beta) \quad (5.10)$$

mit $\beta = (\beta_1^T, \dots, \beta_p^T)^T \in \mathbb{R}^q$ modelliert mit Matrizen $Z_1(x) \in \mathbb{R}^{q \times p}$ (siehe (5.9)), zu schätzendem Parametervektor $\beta \in \mathbb{R}^q$ und einer Linkfunktion $g : \eta \in \mathbb{R}^p \rightarrow g(\eta) \in \mathbb{R}^p$. Die Kovarianzmatrizen werden über die Matrix-Logarithmus-Parametrisierung

$$A(x) = \log(\Sigma(x))$$

modelliert. Dabei fungiert der Logarithmus als Linkfunktion. Die reellen symmetrischen Matrizen $A(x)$ werden wiederum mit fest gewählten Matrizen $U_j (j = 1, \dots, g)$ und einem Parametervektor $\alpha \in \mathbb{R}^g$ wie folgt parametrisiert

$$A(x) = \sum_{j=1}^g [\alpha]_j \cdot U_j(x). \quad (5.11)$$

Soll beispielsweise $A(x) = \begin{pmatrix} [\alpha]_1 & [\alpha]_2 + [\alpha]_3 \cdot x \\ [\alpha]_2 + [\alpha]_3 \cdot x & [\alpha]_4 \end{pmatrix}$ in der Form wie in Gleichung (5.11) dargestellt werden, so bieten sich zum Beispiel die Matrizen

$$U_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, U_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, U_3 = \begin{pmatrix} 0 & x \\ x & 0 \end{pmatrix}, U_4 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

an.

Das vollständige Modell lautet nun

$$\begin{aligned} \mu(x) &= g^{-1}(Z_1(x)^T \beta) \\ A(x) &= \log(\Sigma(x)) = \sum_{j=1}^g [\alpha]_j \cdot U_j(x). \end{aligned} \quad (5.12)$$

Die Schätzung der Parametervektoren $\beta \in \mathbb{R}^q$ und $\alpha \in \mathbb{R}^g$ erfolgt mithilfe der Maximum-Likelihood-Methode. Nach Einsetzen von (5.12) lautet die zweifache logarithmierte Li-

likelihoodfunktion mit (5.8) unter Annahme der multivariaten Normalverteilung

$$2l(\beta, \alpha | Y_1, \dots, Y_n) = -np \log(2\pi) - n \sum_{j=1}^g [\alpha]_j \text{Spur} \left(\sum_{i=1}^n \frac{U_j(x_i)}{n} \right) - \sum_{i=1}^n \text{Spur} \left(B_i \exp(-A(x_i)) \right). \quad (5.13)$$

Hier ist $B_i = \left(Y_i - g^{-1}(Z_1(x_i)^T \beta) \right) \left(Y_i - g^{-1}(Z_1(x_i)^T \beta) \right)^T$. Gesucht ist eine Nullstelle der Ableitung der Likelihoodfunktion $L(\beta, \alpha)$ bzw. der Log-Likelihoodfunktion (5.13).

Im folgenden wird ein Algorithmus zur Berechnung dieser Nullstelle hergeleitet.

Dazu wird im Folgenden sowohl der Gradient als auch die Hesse-Matrix der Log-Likelihoodfunktion (5.13) bezüglich β bestimmt. Des Weiteren werden die partiellen Ableitungen, wie in Chiu et al. (1996), bezüglich α vorgestellt. Abschließend wird ein Algorithmus vorgestellt, der den Newton-Algorithmus (siehe A.1) zur Bestimmung einer Lösung α von $\frac{\partial}{\partial \alpha} l(\beta, \alpha) = 0$ für festes β sowie einer Lösung β von $\frac{\partial}{\partial \beta} l(\beta, \alpha) = 0$ für festes α abwechselnd unter Hinzunahme einer Schrittweitenstrategie verwendet.

Bestimmung des Gradienten und der Hessematrix bezüglich β für festes α

Zunächst wird der Gradient $\frac{\partial}{\partial \beta} l(\beta, \alpha)$ hergeleitet. Dazu wird

$$\begin{aligned} \tilde{f}_i(\beta) &:= \text{Spur}(B_i \exp(-A(x_i))) \\ &= \left(Y_i - g^{-1}(Z_1(x_i)^T \beta) \right)^T \exp(-A(x_i)) \left(Y_i - g^{-1}(Z_1(x_i)^T \beta) \right) \end{aligned} \quad (5.14)$$

für $i = 1, \dots, n$ gesetzt. Mit diesen Bezeichnungen folgt

$$\frac{\partial}{\partial \beta} l(\beta, \alpha) = - \sum_{i=1}^n \frac{\partial}{\partial \beta} \tilde{f}_i(\beta) \quad \text{und} \quad \frac{\partial^2}{\partial \beta \partial \beta^T} l(\beta, \alpha) = - \sum_{i=1}^n \frac{\partial^2}{\partial \beta \partial \beta^T} \tilde{f}_i(\beta).$$

Mit $\tilde{A}(x_i) := \exp(-A(x_i))$ und nach Ausmultiplizieren wird (5.14) zu

$$\tilde{f}_i(\beta) := Y_i^T \tilde{A}(x_i) Y_i - 2g^{-1}(Z_1(x_i)^T \beta)^T \tilde{A}(x_i) Y_i + g^{-1}(Z_1(x_i)^T \beta)^T \tilde{A}(x_i) g^{-1}(Z_1(x_i)^T \beta).$$

Mithilfe der Kettenregel und der Produktregel ergibt sich der Gradient zu

$$\frac{\partial}{\partial \beta} \tilde{f}_i(\beta) = -2Z_1(x_i) \frac{\partial}{\partial \eta} g^{-1}(\eta) \Big|_{Z_1(x_i)^T \beta} \tilde{A}(x_i) \left(Y_i - g^{-1}(Z_1(x_i)^T \beta) \right). \quad (5.15)$$

Damit ergibt sich der r -te Eintrag zu

$$\begin{aligned} \frac{\partial}{\partial [\beta]_r} \tilde{f}_i(\beta) &= -2 \sum_{k=1}^p \sum_{l=1}^p [Z_1(x_i)]_{r,l} \frac{\partial}{\partial \eta_l} g_l^{-1}(\eta) \Big|_{[Z_1(x_i)]_{l,\beta}^T} \\ &\quad \cdot [\tilde{A}(x_i)]_{l,k} \left[Y_i - g^{-1}(Z_1(x_i)^T \beta) \right]_k. \end{aligned} \quad (5.16)$$

Hier bezeichnet $[A]_{k,l}$ den (k, l) -ten Eintrag einer Matrix A und $[a]_k$ meint den k -ten Eintrag eines Vektors a . Leitet man nun (5.16) nach $[\beta]_{r'}$ für $r' \in \{1, \dots, p\}$ ab, so ergibt sich der (r, r') -Eintrag der Hessematrix $\frac{\partial^2}{\partial \beta \partial \beta^T} \tilde{f}_i(\beta)$ durch erneute Anwendung der Produkt- und Kettenregel zu

$$\begin{aligned} \frac{\partial^2}{\partial [\beta]_r \partial [\beta]_{r'}} \tilde{f}_i(\beta) &= -2 \sum_{k=1}^p \sum_{l=1}^p \left\{ [Z_1(x_i)]_{r,l} \right. \\ &\quad \cdot \frac{\partial^2}{\partial^2 \eta} g_l^{-1}(\eta) \Big|_{[Z_1(x_i)]_{l,\beta}^T} [Z_1(x_i)]_{r',l} [\tilde{A}(x_i)]_{l,k} \left[Y_i - g^{-1}(Z_1(x_i)^T \beta) \right]_k \\ &\quad \left. - [Z_1(x_i)]_{r,l} \frac{\partial}{\partial \eta_l} g_l^{-1}(\eta) \Big|_{[Z_1(x_i)]_{l,\beta}^T} \frac{\partial}{\partial \eta_k} g_k^{-1}(\eta_k) \Big|_{[Z_1(x_i)]_{k,\beta}^T} [Z_1(x_i)]_{r',k} [\tilde{A}(x_i)]_{l,k} \right\}. \end{aligned}$$

Zusammenfassend ergibt sich der Gradient und der (r, r') -Eintrag der Hesse-Matrix der Log-Likelihoodfunktion zu

$$\frac{\partial}{\partial \beta} l(\beta, \alpha) = \sum_{i=1}^n -2Z_1(x_i) \frac{\partial}{\partial \eta} g(\eta) \Big|_{Z_1(x_i)^T \beta} \tilde{A}(x_i) \left(Y_i - g^{-1}(Z_1(x_i)^T \beta) \right). \quad (5.17)$$

und

$$\begin{aligned} \frac{\partial^2}{\partial [\beta]_r \partial [\beta]_{r'}} l(\beta, \alpha) &= \sum_{i=1}^n \left(-2 \sum_{k=1}^p \sum_{l=1}^p \left\{ [Z_1(x_i)]_{r,l} \frac{\partial^2}{\partial^2 \eta} g_l^{-1}(\eta) \Big|_{[Z_1(x_i)]_{l,\beta}^T} \right. \right. \\ &\quad \cdot [Z_1(x_i)]_{r',l} [\tilde{A}(x_i)]_{l,k} \left[Y_i - g^{-1}(Z_1(x_i)^T \beta) \right]_k \\ &\quad \left. - [Z_1(x_i)]_{r,l} \frac{\partial}{\partial \eta_l} g_l^{-1}(\eta) \Big|_{[Z_1(x_i)]_{l,\beta}^T} \right. \\ &\quad \left. \cdot \frac{\partial}{\partial \eta_k} g_k^{-1}(\eta_k) \Big|_{[Z_1(x_i)]_{k,\beta}^T} [Z_1(x_i)]_{r',k} [\tilde{A}(x_i)]_{l,k} \right\}. \end{aligned} \quad (5.18)$$

Damit kann die Nullstelle der Log-Likelihoodfunktion bezüglich β für festes α mithilfe des Newton-Verfahrens bestimmt werden.

Bemerkung 5.1 Für die Identitätsabbildung $g = id$, d.h. $g(\eta) = \eta$, ergeben sich die partiellen Ableitungen von Gleichung (5.13) bezüglich β leicht zu

$$\begin{aligned}\frac{\partial l(\beta, \alpha)}{\partial \beta} &= -2 \sum_{i=1}^n Z_1(x_i) \exp(-A(x_i)) (Y_i - Z_1(x_i)^T \beta) \\ \frac{\partial l(\beta, \alpha)}{\partial \beta \partial \beta^T} &= -2 \sum_{i=1}^n Z_1(x_i) \exp(-A(x_i)) Z_1(x_i)^T.\end{aligned}$$

Für festes α kann hier die Lösung von

$$\frac{\partial l(\beta, \alpha)}{\partial \beta} = 0$$

explizit bestimmt werden und ergibt

$$\hat{\beta} = \left\{ \sum_{i=1}^n Z_1(x_i) \exp(-A(x_i)) Z_1(x_i)^T \right\}^{-1} \sum_{i=1}^n Z_1(x_i) \exp(-A(x_i)) Y_i.$$

Bestimmung des Gradienten und der Hessematrix bezüglich α für festes β

Um die partiellen Ableitungen der Log-Likelihoodfunktion (5.13) bezüglich α in einem bestimmten Punkt $\tilde{\alpha}$ zu bestimmen, verwenden Chiu et al. (1996) einen Ansatz zur Lösung einer sogenannten Volterra-Integral-Gleichung und erhalten so eine Approximation zweiter Ordnung für die Likelihoodfunktion (5.13) (Bellman, 1970; Leonard und Hsu, 1992; Chiu et al., 1996) um den Punkt $\tilde{\alpha}$. Sei $X_i(t) = \exp(-A(x_i)t)$. Die Funktion $X_i(t) = \exp(-\tilde{A}(x_i) - (A(x_i) - \tilde{A}(x_i))t)$ mit $\tilde{A}(x_i) = \sum_j \tilde{\alpha}_j U_j(x_i)$ erfüllt die sogenannte Volterra-Integral-Gleichung

$$X_i(t) = \exp(\tilde{A}(x_i))^{-t} - \int_0^t \exp(\tilde{A}(x_i))^{t-s} (A(x_i) - \tilde{A}(x_i)) X_i(s) ds. \quad (5.19)$$

Für $t = 1$ ergibt sich $X_i(1) = \exp\left(-\tilde{A}(x_i) - (A(x_i) - \tilde{A}(x_i)) \cdot 1\right) = \exp(-A(x_i) \cdot 1)$. Wiederholtes Einsetzen von $X_i(t)$ in die rechte Seite der Gleichung (5.19) und Einsetzen des entstehenden Ausdrucks in die Likelihoodfunktion (5.13) ergibt eine Taylor-Entwicklung der Likelihoodfunktion. Eine anschließende Auswertung an der Stelle $t = 1$ liefert damit eine Näherung von $\exp(-A(x_i))$. Der Vergleich mit den Koeffizienten der Taylor-Entwicklung kombiniert mit dem Identitätssatz für Potenzreihen (Königsberger, 2004,

Seite 33) ergibt nun Ausdrücke für die ersten und zweiten partiellen Ableitungen wie folgt (Chiu et al., 1996). Mit

$$r(\beta, \alpha)_i^j = \int_0^1 \text{Spur} \left[\exp(-A(x_i)s) B_i \exp(-A(x_i)(1-s)) U_j(x_i) \right] ds \quad (5.20)$$

und

$$s(\beta, \alpha)_{jk}^i = \int_0^1 \int_0^s \text{Spur} \left[\exp(-A(x_i)u) B_i \exp \left\{ -A(x_i)(1-s) U_j(x_i) \right. \right. \\ \left. \left. \exp(-A(x_i)(s-u)) U_k(x_i) \right\} \right] du ds \quad (5.21)$$

für $j, k = 1, \dots, p$ gilt

$$\frac{\partial l(\beta, \alpha)}{\partial [\alpha]_j} = -n \text{Spur} \left(\sum_{i=1}^n \frac{U_j(x_i)}{n} \right) + \sum_{i=1}^n r(\beta, \alpha)_j^i \quad (5.22)$$

$$\frac{\partial l(\beta, \alpha)}{\partial [\alpha]_j \partial [\alpha]_k} = - \sum_{i=1}^n s(\beta, \alpha)_{jk}^i. \quad (5.23)$$

Bemerkung 5.2 Chiu et al. (1996) zeigen zudem unter gewissen Regularitätsbedingungen Konsistenz und asymptotische Normalität der Maximum-Likelihood-Schätzer $\hat{\beta}$ und $\hat{\alpha}$ für $g = id$. Auf diese Weise bekommen Chiu et al. (1996) einen Ausdruck für die asymptotische Varianz von $\hat{\beta}$ und $\hat{\alpha}$.

Mithilfe der Spektralzerlegung der Matrix $\exp(A(x_i))$ lassen sich die Integrale (5.20) und (5.21) analytisch lösen. Für festes β kann anschließend eine Lösung α unter Verwendung des Newton-Algorithmus (siehe A.1) ermittelt werden.

Sei nun $\exp(A(x_i)) = T_i D_i T_i^T$ die Spektralzerlegung der Matrix $\exp(A(x_i))$ mit einer orthonormalen Matrix $T_i \in \mathbb{R}^{p \times p}$ und einer Diagonalmatrix $D_i \in \mathbb{R}^{p \times p}$ mit Diagonaleinträgen $d_i^r \in \mathbb{R}$ ($r = 1, \dots, p$) für $i = 1, \dots, n$. Damit ergibt sich mit Vertauschung der Produktreihenfolge von Matrizen innerhalb der Spur

$$r(\beta, \alpha)_j^i = \int_0^1 \sum_{k=1}^p \text{Spur} \left\{ D_i^s T_i^T B_i T_i D_i^{s-1} T_i^T U_j(x_i) T_i \right\} ds. \quad (5.24)$$

Bezeichne weiterhin mit u_{ij}^{kw} das (k, w) -Element von $T_i^T U_j(x_i) T_i$ und mit b_i^{kw} (k, w) -Element von $T_i^T B_i T_i$. Mit diesen Bezeichnungen ergibt sich (5.24) zu

$$r(\beta, \alpha)_j^i = \sum_{k=1}^p \sum_{w=1}^p u_{ij}^{wk} b_i^{ww} \int_0^1 (d_i^w)^{s-1} (d_i^k)^{-s} ds. \quad (5.25)$$

Wegen

$$\begin{aligned} \int_0^1 (d_i^w)^{s-1} (d_i^k)^{-s} ds &= \frac{1}{d_i^w} \left(\frac{1}{\ln\left(\frac{d_i^w}{d_i^k}\right)} \left(e^{s \ln\left(\frac{d_i^w}{d_i^k}\right)} \right) \right) \Big|_0^1 \\ &= \frac{1}{d_i^w} \left(\frac{1}{\ln\left(\frac{d_i^w}{d_i^k}\right)} \left(\frac{d_i^w}{d_i^k} - 1 \right) \right) \end{aligned}$$

ergibt sich (5.25) und damit die analytische Lösung von (5.20) schließlich zu

$$r(\beta, \alpha)_j^i = \sum_{k=1}^p \sum_{w=1}^p u_{ij}^{wk} b_i^{ww} (d_i^w)^{-1} \frac{1}{\ln\left(\frac{d_i^w}{d_i^k}\right)} \left(\frac{d_i^w}{d_i^k} - 1 \right). \quad (5.26)$$

Mit denselben oben genannten Bezeichnungen und ähnlichen Argumenten erhält man

$$\begin{aligned} s(\beta, \alpha)_{jk}^i &= \sum_{m=1}^p \sum_{z=1}^p \sum_{r=1}^p b_i^{rm} u_{ij}^{mz} (d_i^r)^{-1} \frac{1}{\ln\left(\frac{d_i^z}{d_i^m}\right)} \left\{ \frac{1}{\ln\left(\frac{d_i^r}{d_i^m}\right)} \left(\frac{d_i^r}{d_i^m} - 1 \right) \right. \\ &\quad \left. - \frac{1}{\ln\left(\frac{d_i^r}{d_i^z}\right)} \left(\frac{d_i^r}{d_i^z} - 1 \right) \right\}. \end{aligned} \quad (5.27)$$

Numerische Umsetzung

Bei der Spektralzerlegung kann es vorkommen, dass Diagonalelemente von D identisch sind. Des Weiteren sind bei der Summation in (5.26) (5.27) für zwei oder drei gleiche Indizes drei der entsprechenden Eigenwerte gleich. Dies führt zu Schwierigkeiten, da beispielsweise in (5.26) für $d_i^w = d_i^k$ der Ausdruck $\ln\left(\frac{d_i^w}{d_i^k}\right)$ Null ergibt und damit divergiert der Kehrwert $\frac{1}{\ln\left(\frac{d_i^w}{d_i^k}\right)}$ gegen Unendlich. Daher müssen diese Fälle zunächst untersucht werden und bei Existenz durch entsprechende Grenzwerte ersetzt werden, damit das Verfahren implementierbar ist. Dazu wird folgendes Lemma 5.3 benötigt.

Lemma 5.3 *Es gilt:*

$$r(\beta, \alpha)_j^i = \left\{ \begin{array}{ll} \sum_{k=1}^p \sum_{w=1}^p u_{ij}^{wk} b_i^{ww} (d_i^w)^{-1} \frac{1}{\ln\left(\frac{d_i^w}{d_i^k}\right)} \left(\frac{d_i^w}{d_i^k} - 1 \right), & \text{für } d_i^w \neq d_i^k, \\ \sum_{k=1}^p \sum_{w=1}^p u_{ij}^{wk} b_i^{ww} (d_i^w)^{-1}, & \text{für } d_i^w = d_i^k. \end{array} \right\} \quad (5.28)$$

$$s(\beta, \alpha)_{jk}^i = \left\{ \begin{array}{l} \sum_{m=1}^p \sum_{z=1}^p \sum_{r=1}^p b_i^{rm} u_{ij}^{mz} (d_i^r)^{-1} \frac{1}{\ln\left(\frac{d_i^z}{d_i^m}\right)} \left\{ \frac{1}{\ln\left(\frac{d_i^r}{d_i^m}\right)} \left(\frac{d_i^r}{d_i^m} - 1\right) \right. \\ \left. - \frac{1}{\ln\left(\frac{d_i^r}{d_i^z}\right)} \left(\frac{d_i^r}{d_i^z} - 1\right) \right\}, \\ \text{für } d_i^r \neq d_i^z \neq d_i^m, \\ \\ \sum_{m=1}^p \sum_{z=1}^p \sum_{r=1}^p b_i^{rm} u_{ij}^{mz} (d_i^r)^{-1} \frac{1}{\ln\left(\frac{d_i^z}{d_i^m}\right)} \left\{ \frac{1}{\ln\left(\frac{d_i^z}{d_i^m}\right)} \right. \\ \left. \cdot \left(\frac{d_i^r}{d_i^m} - 1\right) - 1 \right\}, \\ \text{für } d_i^r = d_i^z \neq d_i^m, \\ \\ \sum_{m=1}^p \sum_{z=1}^p \sum_{r=1}^p b_i^{rm} u_{ij}^{mz} (d_i^r)^{-1} \frac{1 - \frac{d_i^r}{d_i^z} + \frac{d_i^r}{d_i^z} \ln\left(\frac{d_i^r}{d_i^z}\right)}{\left(\ln\left(\frac{d_i^r}{d_i^z}\right)\right)^2}, \\ \text{für } d_i^r \neq d_i^z = d_i^m, \\ \\ \sum_{m=1}^p \sum_{z=1}^p \sum_{r=1}^p b_i^{rm} u_{ij}^{mz} (d_i^r)^{-1} \frac{1}{\ln\left(\frac{d_i^z}{d_i^m}\right)} \left\{ 1 - \frac{1}{\frac{d_i^r}{d_i^z}} \left(\frac{d_i^r}{d_i^z} - 1\right) \right\}, \\ \text{für } d_i^r = d_i^m \neq d_i^z, \\ \\ \sum_{m=1}^p \sum_{z=1}^p \sum_{r=1}^p \frac{1}{2} b_i^{rm} u_{ij}^{mz} (d_i^r)^{-1} \\ \text{für } d_i^r = d_i^z = d_i^m. \end{array} \right. \quad (5.29)$$

Beweis:

Mit der Anwendung der Regeln von l'Hospital (Königsberger, 2004) folgt

$$\lim_{d_i^w \rightarrow d_i^k} \frac{d_i^w}{d_i^k} - 1 \stackrel{x := \frac{d_i^w}{d_i^k}}{=} \lim_{x \rightarrow 1} \frac{x - 1}{\ln(x)} = \lim_{x \rightarrow 1} \frac{1}{\frac{1}{x}} = \lim_{x \rightarrow 1} x = 1.$$

Damit folgt (5.28).

Der zweite und vierte Teil von (5.29) ergibt sich beim Beweis von (5.28) durch einfache Anwendung der Regeln von l'Hospital. Für den dritten Teil von (5.29) wird $x := d_i^m$,

$y := d_i^z$ und $a := d_i^r$ gesetzt. Damit ergibt sich.

$$\lim_{x \rightarrow y} \frac{\frac{1}{\ln\left(\frac{a}{y}\right)}\left(\frac{a}{y} - 1\right) - \frac{1}{\ln\left(\frac{a}{x}\right)}\left(\frac{a}{x} - 1\right)}{\ln\left(\frac{x}{y}\right)} = \lim_{x \rightarrow y} \frac{\ln\left(\frac{a}{x}\right)\left(\frac{a}{y} - 1\right) - \ln\left(\frac{a}{y}\right)\left(\frac{a}{x} - 1\right)}{\ln\left(\frac{a}{y}\right)\ln\left(\frac{x}{y}\right)\ln\left(\frac{a}{x}\right)}.$$

Anwendung der Regeln von l'Hospital ergibt nach Kürzung

$$\begin{aligned} \lim_{x \rightarrow y} \frac{\ln\left(\frac{a}{x}\right)\left(\frac{a}{y} - 1\right) - \ln\left(\frac{a}{y}\right)\left(\frac{a}{x} - 1\right)}{\ln\left(\frac{a}{y}\right)\ln\left(\frac{x}{y}\right)\ln\left(\frac{a}{x}\right)} &= \lim_{x \rightarrow y} \frac{\frac{1}{x} - \frac{a}{xy} + \frac{a}{x^2}\ln\left(\frac{a}{y}\right)}{\frac{1}{x}\ln\left(\frac{a}{y}\right)\ln\left(\frac{a}{x}\right) - \frac{1}{x}\ln\left(\frac{a}{y}\right)\ln\left(\frac{x}{y}\right)} \\ &= \frac{1 - \frac{a}{y} + \frac{a}{y}\ln\left(\frac{a}{y}\right)}{\left(\ln\left(\frac{a}{y}\right)\right)^2}. \end{aligned}$$

Damit folgt der dritte Teil von (5.29).

Für den vierten Teil von (5.29) setze $x := \frac{d_i^z}{d_i^m}$ und $y := \frac{d_i^r}{d_i^m}$. Auf diese Weise erhält man das Ergebnis durch Betrachtung des Grenzwerts bezüglich $(x, y) \rightarrow (1, 1)$. Mit den Bezeichnungen folgt $\frac{y}{x} = \frac{d_i^r}{d_i^z}$. Es gilt nach Multiplikation mit $\frac{x-1}{x-1} (= 1)$ und nach Umsortierung

$$\begin{aligned} &\lim_{(x,y) \rightarrow (1,1)} \frac{\frac{1}{\ln(y)}(y-1) - \frac{1}{\ln\left(\frac{y}{x}\right)}\left(\frac{y}{x} - 1\right)}{\ln(x)} \frac{x-1}{x-1} \\ &= \lim_{(x,y) \rightarrow (1,1)} \underbrace{\left(\frac{\ln(x)}{x-1}\right)^{-1}}_{:= (\star)} (-1) \cdot \underbrace{\frac{\frac{1}{\ln\left(\frac{y}{x}\right)}\left(\frac{y}{x} - 1\right) - \frac{1}{\ln(y)}(y-1)}{x-1}}_{:= (\star\star)} \\ &=: (\star\star\star) \end{aligned}$$

Bei Existenz ist der Grenzwert eines Produkts das Produkt der einzelnen Grenzwerte.

Daher werden die Grenzwerte für (\star) und $(\star\star)$ näher untersucht. Für (\star) gilt $\lim_{(x,y) \rightarrow (1,1)} (\star) =$

$\left(\left(\frac{\partial}{\partial x} \ln(x)\right)\Big|_{x=1}\right)^{-1} = (x)\Big|_{x=1} = 1$. Setzt man nun $f(x, y) := \frac{1}{\ln\left(\frac{y}{x}\right)}\left(\frac{y}{x} - 1\right)$, dann gilt

$$\lim_{(x,y) \rightarrow (1,1)} (\star\star) = \lim_{(x,y) \rightarrow (1,1)} \frac{f(x, y) - f(1, y)}{x-1} = \left(\frac{\partial}{\partial x} f(x, y)\right)\Big|_{(x,y)=(1,1)}$$

Die partielle Ableitung ergibt sich zu

$$\frac{\partial}{\partial x} f(x, y) = \frac{\frac{1}{x} \frac{y}{x} - 1 - \frac{y}{x} \ln\left(\frac{y}{x}\right)}{\left(\ln\left(\frac{y}{x}\right)\right)^2} \stackrel{z := \frac{y}{x}}{=} \frac{1}{x} \frac{z - 1 - z \ln(z)}{(\ln(z))^2}$$

Damit ergibt sich mit $z := \frac{y}{x}$ und Anwendung der Regeln von l'Hospital

$$\begin{aligned}
\left. \left(\frac{\partial}{\partial x} f(x, y) \right) \right|_{(x,y)=(1,1)} &= \lim_{(x,y) \rightarrow (1,1)} \frac{\partial}{\partial x} f(x, y) \\
&= \lim_{(x,z) \rightarrow (1,1)} \frac{1}{x} \cdot \frac{z - 1 - z \ln(z)}{(\ln(z))^2} \\
&= \lim_{z \rightarrow 1} \frac{1 - 1 - \ln(z)}{2 \ln(z) \frac{1}{z}} \\
&= \lim_{z \rightarrow 1} \frac{-z \ln(z)}{2 \ln(z)} = \lim_{z \rightarrow 1} \frac{-z}{2} = -\frac{1}{2}.
\end{aligned}$$

Insgesamt ergibt sich wegen $\lim_{(x,y) \rightarrow (1,1)} (\star) = 1$ und $\lim_{(x,y) \rightarrow (1,1)} (\star\star) = -\frac{1}{2}$ die Behauptung

$$\frac{1}{2} = 1 \cdot (-1) \cdot \left(-\frac{1}{2}\right) = \lim_{(x,y) \rightarrow (1,1)} (\star) \cdot (-1) \cdot \lim_{(x,y) \rightarrow (1,1)} (\star\star) = \lim_{(x,y) \rightarrow (1,1)} (\star\star\star).$$

Insgesamt folgt die Behauptung. □

Mit den Ausdrücken (5.28) und (5.29) lässt sich der Gradient (5.22) sowie die Hessematrix (5.23) einfach und schnell berechnen. Nun kann das Newton-Verfahren zum Einsatz kommen.

Die Konvergenz des Newton-Verfahrens ist im Allgemeinen nur lokal gegeben und daher hängt der Erfolg des Verfahrens von einer geeigneten Wahl der Startwerte ab. Daher ist der Algorithmus in Chiu et al. (1996) in der Praxis nicht direkt umsetzbar, da häufig eine Divergenz des Verfahrens eintritt. Aufgrund dessen wird in dieser Arbeit eine Schrittweitenstrategie zur Bestimmung der Koeffizientenvektoren (für $g \neq id$) angewandt, um das Konvergenzverhalten des Verfahrens zu verbessern. Es ist nicht garantiert, dass die Iterationsschritte $\alpha^{(m+1)}$ ohne Schrittweitensteuerung immer einen garantierten Abstieg zur Folge haben. Es wurde in ersten Tests vielfach beobachtet, dass die Folge divergieren kann, da die Newton-Schritte unter Umständen zu groß sind. Daher wird eine Schrittweite $t \in (0, 1)$ gesucht, die zu einem garantierten Abstieg führt.

Der Algorithmus zur Bestimmung einer Maximum-Likelihood-Schätzung der Koeffizientenvektoren $(\beta, \alpha) \in \mathbb{R}^{q+w}$ wird nun wie folgt zusammengefasst.

Algorithmus 5.4 (Anpassung des Matrix-Logarithmus-Modells)

1. Lege Konstante $\delta > 0$ zur Schrittweitensteuerung, eine Abbruchkonstante $\epsilon > 0$ sowie Startvektoren $\alpha^{(0)} \in \mathbb{R}^w$ und $\beta^{(0)} \in \mathbb{R}^q$ fest. Setze $s = 0$.

2. Schritt $(s+1)$:

(a) Berechnung von $\beta^{(s+1)}$.

(i) Setze $\beta_s^{(0)} := \beta^{(s)}$ und setze $m = 0$.

(ii) Schritt $(m+1)$:

Berechne

$$d_\beta^{(m+1)} = \left\{ \frac{\partial l(\beta, \alpha_s)}{\partial \beta \partial \beta^T} \right\}^{-1} \bigg|_{\beta_s^{(m)}} \frac{\partial l(\beta, \alpha_s)}{\partial \beta} \bigg|_{\beta_s^{(m)}}.$$

Berechne nun für $k = 0, 1, 2, 3, \dots$

$$\beta_s^{(m+1)} = \beta_s^{(m)} - \delta^k d_\beta^{m+1}.$$

bis zu einem $\hat{k} \in \mathbb{N}$, so dass gilt $l(\beta_s^{(m+1)}, \alpha_s) < l(\beta^{(s)}, \alpha_s)$.

(iii) Falls gilt $\left\| \frac{\partial l(\beta, \alpha_s)}{\partial \beta} \bigg|_{\beta = \beta_s^{(m+1)}} \right\| < \epsilon$, so setze $\beta^{(s+1)} = \beta_s^{(m+1)}$ und gehe zu 2(b)(i). Andernfalls gehe zu 2(a)(ii).

(b) Berechnung von $\alpha^{(s+1)}$.

(i) Setze $\alpha_s^{(0)} := \alpha^{(s)}$ und setze $m = 0$.

(ii) Schritt $(m+1)$:

Berechne

$$d_\alpha^{(m+1)} = \left\{ \frac{\partial l(\beta^{(s+1)}, \alpha)}{\partial \alpha \partial \alpha^T} \right\}^{-1} \bigg|_{\alpha_s^{(m)}} \frac{\partial l(\beta^{(s+1)}, \alpha)}{\partial \alpha} \bigg|_{\alpha_s^{(m)}}.$$

Berechne nun für $k = 0, 1, 2, 3, \dots$

$$\alpha_s^{(m+1)} = \alpha_s^{(m)} - \delta^k d_\alpha^{(m+1)}.$$

bis zu einem $\hat{k} \in \mathbb{N}$, so dass gilt $l(\beta^{(s+1)}, \alpha_s^{(m+1)}) < l(\beta^{(s+1)}, \alpha_s^{(m)})$.

(iii) Falls gilt $\left\| \frac{\partial l(\beta^{(s+1)}, \alpha)}{\partial \alpha} \Big|_{\alpha = \alpha_s^{(m+1)}} \right\| < \epsilon$, so setze $\alpha^{(s+1)} = \alpha_s^{(m+1)}$ und gehe zu 2(c). Andernfalls gehe zu 2(b)(ii).

(c) Falls gilt $\left\| \frac{\partial l(\beta, \alpha)}{\partial (\beta, \alpha)} \Big|_{(\beta, \alpha) = (\beta^{(s+1)}, \alpha^{(s+1)})} \right\| < \epsilon$ und falls gilt $\|\alpha^{(s+1)} - \alpha^{(s)}\| < \epsilon$, so gehe zu 3. Andernfalls setze $s = s + 1$ und gehe zu 2(a)(i).

3. Abbruch! Setze $(\hat{\beta}, \hat{\alpha}) = (\beta^{(s+1)}, \alpha^{(s+1)})$.

5.3. Schätzer basierend auf dem Kovarianzregressionsmodell mit einer allgemeinen Linkfunktion

Hoff und Niu (2012) stellen ein Kovarianzregressionsmodell vor, das die gleichzeitige Modellierung des Erwartungswertvektors und der Kovarianzmatrix in Abhängigkeit von Kovariablen erlaubt. Allerdings umfasst das Modell nicht die Möglichkeit, Designvektoren $f_i(x) \in \mathbb{R}^{q_i}$ unterschiedlicher Länge für die unterschiedlichen Zielvariablen Y_i , $i = 1, \dots, p$, zu verwenden. Damit können einzelne Designvariablen im Zuge einer Modellselktion nicht direkt für eine einzelne Zielvariable entfernt oder hinzugefügt werden. Daher wird das Modell in diesem Abschnitt mithilfe einer alternativen Parametrisierung durch Verwendung der Matrix $Z_1(x)$ (siehe (5.9)) für den Erwartungswertvektor und $Z_2(x)$ zur Modellierung der Kovarianzmatrix mit

$$Z_2(x) = \begin{pmatrix} \tilde{f}_1(x) & 0_{w_1} & 0_{w_1} & \dots & 0_{w_1} & 0_{w_1} \\ 0_{w_2} & \tilde{f}_2(x) & 0_{w_2} & \dots & 0_{w_2} & 0_{w_2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0_{w_{p-1}} & 0_{w_{p-1}} & 0_{w_{p-1}} & \dots & \tilde{f}_{p-1}(x) & 0_{w_{p-1}} \\ 0_{w_p} & 0_{w_p} & 0_{w_p} & \dots & 0_{w_p} & \tilde{f}_p(x) \end{pmatrix} \in \mathbb{R}^{p \times w} \quad (5.30)$$

mit Designvektoren $\tilde{f}_j(x) \in \mathbb{R}^{w_j}$ für $j \in \{1, \dots, p\}$ erweitert. Beide Matrizen lassen Designvektoren unterschiedlicher Länge zu. Damit können Variablen für einzelne Zielgrößen aus dem Modell entfernt oder hinzugefügt werden und unterschiedliche Modelle für jede Zielgröße selektiert werden. Der Erwartungswertvektor wird über die Gleichung (5.10)

modelliert. Das vollständige Modell lautet

$$\begin{aligned}\mu(x) &= g^{-1}(Z_1(x)^T \beta) \\ \Sigma(x) &= \Psi + (Z_2(x)^T \alpha)(Z_2(x)^T \alpha)^T\end{aligned}\tag{5.31}$$

mit einer unbekanntem positiv definiten Matrix $\Psi \in \mathbb{R}^{p \times p}$ und einem weiteren zu schätzenden Koeffizientenvektor $\alpha \in \mathbb{R}^w$ ($i = 1, \dots, n$). Die einzelnen Komponenten des Erwartungswertvektors werden hier über eine Linkfunktion mit einem linearen Modell verknüpft. Das Modell für die Kovarianzmatrix wird als Summe einer konstanten Matrix Ψ und einer symmetrischen Matrix vom Rang 1 vom Typ aa^T modelliert. Die einzelnen Komponenten von a bestehen aus linearen Modellen.

Zu bemerken ist, dass der Koeffizientenvektor α auch bei dieser Parametrisierung bis auf das Vorzeichen identifizierbar ist, wie im folgenden Satz für den Fall einer Matrix $Z_2(x)$ mit $\tilde{f}_j(x) = (1, [x]_1, \dots, [x]_n)$ gezeigt wird. In dieser Arbeit werden lediglich Modelle betrachtet, die nur Haupteffekte in den Kovariablenvektoren $\tilde{f}_j(x)$ enthalten und für diesen Fall zeigt der folgende Satz die Identifizierbarkeit des Kovarianzregressionsmodells.

Satz 5.5 *Sei $x \in \mathbb{R}^k$ und $\Sigma(x) = \Psi + (Z_2(x)^T \alpha)(Z_2(x)^T \alpha)^T$ wie in (5.31) mit einer Matrix $Z_2(x)$ wie in (5.30) gegeben. Seien die Kovariablenvektoren $\tilde{f}_i(x)$ gegeben durch $\tilde{f}_i(x) = (1, [x]_1, \dots, [x]_k)$. Sei eine weitere Matrix $\tilde{\Sigma}(x) = \tilde{\Psi} + (Z_2(x)^T \tilde{\alpha})(Z_2(x)^T \tilde{\alpha})^T$ gegeben mit $\Sigma(x) = \tilde{\Sigma}(x) \forall x \in \mathbb{R}^k$. Dann gilt $\Psi = \tilde{\Psi}$ und $\alpha = \pm \tilde{\alpha}$.*

Beweis:

Für Vektoren $x^{(i)} \in \mathbb{R}^k$ mit

$$x^{(i)} = \begin{cases} [x]_i & , \text{i-ter Eintrag von } x^{(i)} \\ 0 & , \text{sonst} \end{cases}\tag{5.32}$$

gilt

$$Z_2(x^{(i)})^T \alpha = \begin{pmatrix} [\alpha]_1 + [\alpha]_{1+i}[x]_i \\ \vdots \\ [\alpha]_{1+(k+1)(j-1)} + [\alpha]_{1+(k+1)(j-1)+i}[x]_i \\ \vdots \\ [\alpha]_{1+(k+1)(p-1)} + [\alpha]_{1+(k+1)(p-1)+i}[x]_i \end{pmatrix}.$$

Folglich sind die Produkte $Z_2(x^{(i)})^T \alpha$ für alle solche Vektoren $x^{(i)}$ für $i = 1, \dots, k$ von der Struktur her gleich. Sei für ein beliebiges $i \in \{1, \dots, k\}$ ein $x^{(i)}$ entsprechend (5.32) gegeben. Dann ergibt sich für $j = 1, \dots, p$ und $i = 1, \dots, k$

$$\begin{aligned} [\Sigma(x^{(i)})]_{jj} &= [\Psi]_{jj} + ([\alpha]_{1+(k+1)(j-1)} + [\alpha]_{1+(k+1)(j-1)+i}[x]_i)^2 \\ &= [\Psi]_{jj} + [\alpha]_{1+(k+1)(j-1)}^2 + 2[\alpha]_{1+(k+1)(j-1)}[\alpha]_{1+(k+1)(j-1)+i}[x]_i \\ &\quad + [\alpha]_{1+(k+1)(j-1)+i}^2[x]_i^2 \end{aligned} \quad (5.33)$$

und

$$\begin{aligned} [\tilde{\Sigma}(x^{(i)})]_{jj} &= [\tilde{\Psi}]_{jj} + ([\tilde{\alpha}]_{1+(k+1)(j-1)} + [\tilde{\alpha}]_{1+(k+1)(j-1)+i}[x]_i)^2 \\ &= [\tilde{\Psi}]_{jj} + [\tilde{\alpha}]_{1+(k+1)(j-1)}^2 + 2[\tilde{\alpha}]_{1+(k+1)(j-1)}[\tilde{\alpha}]_{1+(k+1)(j-1)+i}[x]_i \\ &\quad + [\tilde{\alpha}]_{1+(k+1)(j-1)+i}^2[x]_i^2 \end{aligned} \quad (5.34)$$

Mit der Wahl von $[x]_i = 0$ ergibt sich

$$[\Psi]_{jj} + [\alpha]_{1+(k+1)(j-1)}^2 = [\tilde{\Psi}]_{jj} + [\tilde{\alpha}]_{1+(k+1)(j-1)}^2. \quad (5.35)$$

Damit folgt

$$\begin{aligned} &2[\tilde{\alpha}]_{1+(k+1)(j-1)}[\tilde{\alpha}]_{1+(k+1)(j-1)+i}[x]_i + [\tilde{\alpha}]_{1+(k+1)(j-1)+i}^2[x]_i^2 \\ &= 2[\alpha]_{1+(k+1)(j-1)}[\alpha]_{1+(k+1)(j-1)+i}[x]_i + [\alpha]_{1+(k+1)(j-1)+i}^2[x]_i^2. \end{aligned} \quad (5.36)$$

Durch die Wahl von $[x]_i = 1$ und $[x]_i = -1$ ergibt sich

$$\begin{aligned} &2[\tilde{\alpha}]_{1+(k+1)(j-1)}[\tilde{\alpha}]_{1+(k+1)(j-1)+i} + [\tilde{\alpha}]_{1+(k+1)(j-1)+i}^2 \\ &= 2[\alpha]_{1+(k+1)(j-1)}[\alpha]_{1+(k+1)(j-1)+i} + [\alpha]_{1+(k+1)(j-1)+i}^2 \end{aligned} \quad (5.37)$$

und

$$\begin{aligned} &-2[\tilde{\alpha}]_{1+(k+1)(j-1)}[\tilde{\alpha}]_{1+(k+1)(j-1)+i} + [\tilde{\alpha}]_{1+(k+1)(j-1)+i}^2 \\ &= -2[\alpha]_{1+(k+1)(j-1)}[\alpha]_{1+(k+1)(j-1)+i} + [\alpha]_{1+(k+1)(j-1)+i}^2. \end{aligned} \quad (5.38)$$

Addition der beiden Gleichungen (5.37) und (5.38) ergibt

$$2[\tilde{\alpha}]_{1+(k+1)(j-1)+i}^2 = 2[\alpha]_{1+(k+1)(j-1)+i}^2 \quad (5.39)$$

und es folgt

$$[\tilde{\alpha}]_{1+(k+1)(j-1)+i} = \pm[\alpha]_{1+(k+1)(j-1)+i}. \quad (5.40)$$

Mit Gleichung (5.37) und der Gleichheit (5.40) ergibt sich für alle $i = 1, \dots, k$ und $j = 1, \dots, p$

$$\begin{aligned} [\tilde{\alpha}]_{1+(k+1)(j-1)+i} &= \pm[\alpha]_{1+(k+1)(j-1)+i} \\ [\tilde{\alpha}]_{1+(k+1)(j-1)} &= \pm[\alpha]_{1+(k+1)(j-1)}. \end{aligned}$$

Zusammenfassend folgt damit $\alpha = \pm\tilde{\alpha}$ und wegen der Gleichheit $\Sigma(x) = \tilde{\Sigma}(x)$ folgt $\Psi = \tilde{\Psi}$. \square

Der Fall, dass die Designvektoren $\tilde{f}_j(x)$ unterschiedliche Längen haben und aus einer Teilmenge der Einträge des Kovariablenvektors x zusammengesetzt sind, ist ein wichtiger Spezialfall des Satzes 5.5.

Im folgenden wird ein Algorithmus zur Schätzung der unbekannt Parameter im Kovarianzregressionsmodell hergeleitet. Die doppelte negative Log-Likelihoodfunktion basierend auf der Normalverteilung lautet für das Kovarianzregressionsmodell

$$\begin{aligned} -2l(\alpha, \beta, \Psi) &= np \log(2\pi) + \sum_{i=1}^n \log |\Psi + (Z_2(x_i)^T \alpha)(Z_2(x_i)^T \alpha)^T| + \\ &\sum_{i=1}^n \left(Y_i - g^{-1}(Z_1(x_i)^T \beta) \right)^T \left(\Psi + (Z_2(x_i)^T \alpha)(Z_2(x_i)^T \alpha)^T \right)^{-1} \left(Y_i - g^{-1}(Z_1(x_i)^T \beta) \right) \end{aligned} \quad (5.41)$$

und die direkte Minimierung von (5.41) ist schwierig, insbesondere wegen der Terme $\log |\Psi + (Z_2(x_i)^T \alpha)(Z_2(x_i)^T \alpha)^T|$ und $\left(\Psi + (Z_2(x_i)^T \alpha)(Z_2(x_i)^T \alpha)^T \right)^{-1}$. Die Maximum-Likelihood-Schätzer können etwas einfacher mithilfe des EM-Algorithmus bestimmt werden, deren Ausgangspunkt folgende Interpretation des Kovarianzregressionsmodells darstellt. Das Kovarianzregressionsmodell ergibt sich aus folgendem Modell

$$Y = g^{-1}(Z_1(x)^T \beta) + (Z_2(x)^T \alpha) \cdot \gamma + \epsilon \quad (5.42)$$

mit zufälligen Effekten γ . Hier ist ϵ ein p -dimensionaler Zufallsvektor und γ eine eindimensionale Zufallsvariable mit den Annahmen $\gamma \sim N(0, 1)$, $\epsilon \sim N(0, \Psi)$ und $E(\gamma \cdot \epsilon) = 0$. Die Berechnung des Erwartungswerts und der Kovarianzmatrix von (5.42) ergibt das Modell (5.31). Die Variablen γ in der Formulierung (5.42) können als fehlende Werte interpretiert werden und auf diese Weise kann das Kovarianzregressionsmodell mithilfe des EM-Algorithmus geschätzt werden. Bezeichne f_X die Dichtefunktion einer Zufallsvariable X bezüglich des Lebesgue-Maßes. Die doppelte negative Log-Likelihoodfunktion basierend auf (5.42) ergibt sich aufgrund der Überlegung

$$\begin{aligned} -2 \log(f_{Y,\gamma}(Y, \gamma|\alpha, \beta, \Psi)) &= -2 \log\left(f_{Y|\gamma}(Y|\gamma, \alpha, \beta, \Psi) f_\gamma(\gamma|\alpha, \beta, \Psi)\right) \\ &= -2 \log\left(f_{Y|\gamma}(Y|\gamma, \alpha, \beta, \Psi)\right) - 2 \log\left(f_\gamma(\gamma|\alpha, \beta, \Psi)\right). \end{aligned} \tag{5.43}$$

Aufgrund der Normalverteilungsannahme folgt

$$\begin{aligned} -2 \log\left(f_{Y|\gamma}(Y|\gamma, \alpha, \beta, \Psi)\right) &= p \log(2\pi) + \log|\Psi| \\ &+ (Y - g^{-1}(Z_1(x)^T \beta) - \gamma Z_2(x)^T \alpha)^T \Psi^{-1} (Y - g^{-1}(Z_1(x)^T \beta) - \gamma Z_2(x)^T \alpha) \end{aligned}$$

und

$$-2 \log\left(f_\gamma(\gamma|\alpha, \beta, \Psi)\right) = \log(2\pi) + \gamma^2$$

und damit insgesamt

$$\begin{aligned} -2 \log\left(f_{Y,\gamma}(Y, \gamma|\alpha, \beta, \Psi)\right) &= p \log(2\pi) + \log|\Psi| \\ &+ (Y - g^{-1}(Z_1(x)^T \beta) - \gamma Z_2(x)^T \alpha)^T \Psi^{-1} (Y - g^{-1}(Z_1(x)^T \beta) - \gamma Z_2(x)^T \alpha) \\ &+ \log(2\pi) + \gamma^2. \end{aligned} \tag{5.44}$$

Schätzung mithilfe des EM-Algorithmus

Die Schätzung erfolgt mithilfe einer Variante des EM-Algorithmus, dem ECM-Algorithmus, (Dempster et al., 1977; McLachlan und Krishnan, 2008) auf Basis von n Stichpro-

benvariablen $Y_1|x_1, \dots, Y_n|x_n$ mit zugehörigen Kovariablenvektoren x_1, \dots, x_n . Der EM-Algorithmus mit einigen Varianten wie dem generalisierten EM-Algorithmus sowie dem ECM-Algorithmus wird in McLachlan und Krishnan (2008) umfassend beschrieben. Der Algorithmus besteht in jeder Iteration aus zwei Schritten, der Berechnung der erwarteten bedingten Likelihoodfunktion (**E-Step**) und der (nicht notwendigerweise vollständigen) Maximierung (oder Minimierung der negativen Log-Likelihoodfunktion) der erwarteten bedingten Log-Likelihoodfunktion (**M-step**). Diese beiden Schritte werden im Folgenden hergeleitet, bevor der vollständige Algorithmus zusammenfassend vorgestellt wird.

1. E-Step

Im E-Step wird zunächst mithilfe von (5.44) im ersten Schritt der bedingte Erwartungswert der doppelten negativen Log-Likelihoodfunktion berechnet und lautet

$$\begin{aligned} \tilde{l}(\alpha, \beta, \Psi) &:= E(-2l(\alpha, \beta, \Psi) | Y_i, x_i, \tilde{\beta}, \tilde{\alpha}, \tilde{\Psi}) \\ &= np \log 2(\pi) + n \log |\Psi| + \\ &\sum_{i=1}^n E \left((Y_i - g^{-1}(Z_1(x_i)^T \beta) - \gamma_i Z_2(x_i)^T \alpha)^T \Psi^{-1} (Y_i - g^{-1}(Z_1(x_i)^T \beta) \right. \\ &\quad \left. - \gamma_i Z_2(x_i)^T \alpha) | Y_i, x_i, \tilde{\beta}, \tilde{\alpha}, \tilde{\Psi} \right) + \sum_{i=1}^n E \left(\log(2\pi) + \gamma_i^2 | Y_i, x_i, \tilde{\beta}, \tilde{\alpha}, \tilde{\Psi} \right). \end{aligned}$$

Mit der Linearität des Erwartungswertes ergibt sich zunächst

$$\begin{aligned} \tilde{l}(\alpha, \beta, \Psi) &= E \left(-2l(\alpha, \beta, \Psi) | Y_i, x_i, \tilde{\beta}, \tilde{\alpha}, \tilde{\Psi} \right) \\ &= n(p+1) \log(2\pi) + \sum_{i=1}^n E \left(\gamma_i^2 | Y_i, x_i, \tilde{\beta}, \tilde{\alpha}, \tilde{\Psi} \right) + n \log |\Psi| + \\ &\sum_{i=1}^n E \left((Y_i - g^{-1}(Z_1(x_i)^T \beta) - \gamma_i Z_2(x_i)^T \alpha)^T \Psi^{-1} (Y_i - g^{-1}(Z_1(x_i)^T \beta) \right. \\ &\quad \left. - \gamma_i Z_2(x_i)^T \alpha) | Y_i, x_i, \tilde{\beta}, \tilde{\alpha}, \tilde{\Psi} \right). \end{aligned}$$

Dazu werden der Erwartungswert und die Varianz der bedingten Zufallsvariable $(\gamma_i | Y_i, x_i, \tilde{\alpha}, \tilde{\beta}, \tilde{\Psi})$ benötigt. Hierfür wird die Verteilung von $((\gamma_i, Y_i) | x_i, \tilde{\alpha}, \tilde{\beta}, \tilde{\Psi})^T$ ge-

braucht. Dieser Zufallsvektor ist lediglich eine lineare Transformation von $((\gamma_i, \epsilon_i) | x_i, \tilde{\alpha}, \tilde{\beta}, \tilde{\Psi})^T$ und ist daher multivariat normalverteilt mit

$$((\gamma_i, Y_i) | x_i, \tilde{\alpha}, \tilde{\beta}, \tilde{\Psi})^T \sim N\left(\begin{pmatrix} 0 \\ g^{-1}(Z_1(x_i)^T \tilde{\beta}) \end{pmatrix}, \begin{pmatrix} 1 & (Z_2(x_i)^T \tilde{\alpha})^T \\ Z_2(x_i)^T \tilde{\alpha} & \tilde{\Psi} + (Z_2(x_i)^T \tilde{\alpha})(Z_2(x_i)^T \tilde{\alpha})^T \end{pmatrix}\right).$$

Folglich ergibt sich mit Anwendung von Satz 2.44 aus Wasserman (2004)

$$(\gamma_i | Y_i, x_i, \tilde{\alpha}, \tilde{\beta}, \tilde{\Psi}) \sim N(m_i, v_i)$$

mit

$$v_i = 1 - (Z_2(x_i)^T \tilde{\alpha})^T \left[\tilde{\Psi} + (Z_2(x_i)^T \tilde{\alpha})(Z_2(x_i)^T \tilde{\alpha})^T \right]^{-1} Z_2(x_i)^T \tilde{\alpha} \text{ und}$$

$$m_i = 0 + (Z_2(x_i)^T \tilde{\alpha})^T \left[\tilde{\Psi} + (Z_2(x_i)^T \tilde{\alpha})(Z_2(x_i)^T \tilde{\alpha})^T \right]^{-1} \left(Y_i - g^{-1}(Z_1(x_i)^T \tilde{\beta}) \right)^T.$$

Durch Anwendung der Sherman-Morrison-Formel $(A + b \cdot b^T)^{-1} = A^{-1} - \frac{A^{-1} b b^T A^{-1}}{1 + b^T A^{-1} b}$ für eine Matrix $A \in \mathbb{R}^{p \times p}$ und einen Vektor $b \in \mathbb{R}^p$ (Hager, 1989) folgt

$$\begin{aligned} v_i &= 1 - (Z_2(x_i)^T \tilde{\alpha})^T \left[\tilde{\Psi}^{-1} - \frac{\tilde{\Psi}^{-1} (Z_2(x_i)^T \tilde{\alpha})(Z_2(x_i)^T \tilde{\alpha})^T \tilde{\Psi}^{-1}}{1 + (Z_2(x_i)^T \tilde{\alpha})^T \tilde{\Psi}^{-1} (Z_2(x_i)^T \tilde{\alpha})} \right] Z_2(x_i)^T \tilde{\alpha} \\ &= 1 - (Z_2(x_i)^T \tilde{\alpha})^T \tilde{\Psi}^{-1} (Z_2(x_i)^T \tilde{\alpha}) \\ &\quad + \frac{(Z_2(x_i)^T \tilde{\alpha})^T \tilde{\Psi}^{-1} (Z_2(x_i)^T \tilde{\alpha})(Z_2(x_i)^T \tilde{\alpha})^T \tilde{\Psi}^{-1} Z_2(x_i)^T \tilde{\alpha}}{1 + (Z_2(x_i)^T \tilde{\alpha})^T \tilde{\Psi}^{-1} (Z_2(x_i)^T \tilde{\alpha})} \\ &= \frac{1^2 - (Z_2(x_i)^T \tilde{\alpha})^T \tilde{\Psi}^{-1} (Z_2(x_i)^T \tilde{\alpha})(Z_2(x_i)^T \tilde{\alpha})^T \tilde{\Psi}^{-1} (Z_2(x_i)^T \tilde{\alpha})}{1 + (Z_2(x_i)^T \tilde{\alpha})^T \tilde{\Psi}^{-1} (Z_2(x_i)^T \tilde{\alpha})} \\ &\quad + \frac{(Z_2(x_i)^T \tilde{\alpha})^T \tilde{\Psi}^{-1} (Z_2(x_i)^T \tilde{\alpha})(Z_2(x_i)^T \tilde{\alpha})^T \tilde{\Psi}^{-1} Z_2(x_i)^T \tilde{\alpha}}{1 + (Z_2(x_i)^T \tilde{\alpha})^T \tilde{\Psi}^{-1} (Z_2(x_i)^T \tilde{\alpha})} \\ &= \frac{1}{1 + (Z_2(x_i)^T \tilde{\alpha})^T \tilde{\Psi}^{-1} (Z_2(x_i)^T \tilde{\alpha})}. \end{aligned}$$

Für die Größe m_i ergibt sich ebenfalls mit Anwendung der Sherman-Morrison-Formel

$$\begin{aligned}
m_i &= 0 + (Z_2(x_i)^T \tilde{\alpha})^T \left[\tilde{\Psi} + (Z_2(x_i)^T \tilde{\alpha}) (Z_2(x_i)^T \tilde{\alpha})^T \right]^{-1} \left(Y_i - g^{-1}(Z_1(x_i)^T \tilde{\beta}) \right)^T \\
&= (Z_2(x_i)^T \tilde{\alpha})^T \left[\tilde{\Psi}^{-1} - \frac{\tilde{\Psi}^{-1} (Z_2(x_i)^T \tilde{\alpha}) (Z_2(x_i)^T \tilde{\alpha})^T \tilde{\Psi}^{-1}}{1 + (Z_2(x_i)^T \tilde{\alpha})^T \tilde{\Psi}^{-1} (Z_2(x_i)^T \tilde{\alpha})} \right] \left(Y_i - g^{-1}(Z_1(x_i)^T \tilde{\beta}) \right)^T \\
&= \left[(Z_2(x_i)^T \tilde{\alpha})^T \tilde{\Psi}^{-1} - \frac{(Z_2(x_i)^T \tilde{\alpha})^T \tilde{\Psi}^{-1} (Z_2(x_i)^T \tilde{\alpha}) (Z_2(x_i)^T \tilde{\alpha})^T \tilde{\Psi}^{-1}}{1 + (Z_2(x_i)^T \tilde{\alpha})^T \tilde{\Psi}^{-1} (Z_2(x_i)^T \tilde{\alpha})} \right] \\
&\quad \cdot \left(Y_i - g^{-1}(Z_1(x_i)^T \tilde{\beta}) \right)^T \\
&= \left[1 - \frac{(Z_2(x_i)^T \tilde{\alpha})^T \tilde{\Psi}^{-1} (Z_2(x_i)^T \tilde{\alpha})}{1 + (Z_2(x_i)^T \tilde{\alpha})^T \tilde{\Psi}^{-1} (Z_2(x_i)^T \tilde{\alpha})} \right] (Z_2(x_i)^T \tilde{\alpha})^T \tilde{\Psi}^{-1} \left(Y_i - g^{-1}(Z_1(x_i)^T \tilde{\beta}) \right)^T \\
&= v_i \left(Y_i - g^{-1}(Z_1(x_i)^T \tilde{\beta}) \right)^T \tilde{\Psi}^{-1} (Z_2(x_i)^T \tilde{\alpha}).
\end{aligned}$$

Damit erhält man insgesamt $(\gamma_i | Y_i, x_i, \tilde{\alpha}, \tilde{\beta}, \tilde{\Psi}) \sim N(m_i, v_i)$ mit

$$\begin{aligned}
v_i &= \left[1 + (Z_2(x_i)^T \tilde{\alpha})^T \tilde{\Psi}^{-1} Z_2(x_i)^T \tilde{\alpha} \right]^{-1} \quad \text{und} \\
m_i &= v_i \left(Y_i - g^{-1}(Z_1(x_i)^T \tilde{\beta}) \right)^T \tilde{\Psi}^{-1} Z_2(x_i)^T \tilde{\alpha}
\end{aligned} \tag{5.45}$$

für $i = 1, \dots, n$. Nun kann die erwartete Log-Likelihoodfunktion gegeben $(Y_i, x_i, \tilde{\alpha}, \tilde{\beta}, \tilde{\Psi})$ berechnet werden wie folgt

$$\begin{aligned}
\tilde{l}(\alpha, \beta, \Psi) &= n(p+1) \log(2\pi) + \sum_{i=1}^n E \left(\gamma_i^2 | Y_i, x_i, \tilde{\beta}, \tilde{\alpha}, \tilde{\Psi} \right) + n \log |\Psi| + \\
&\quad \sum_{i=1}^n E \left((Y_i - g^{-1}(Z_1(x_i)^T \beta) - \gamma_i Z_2(x_i)^T \alpha)^T \Psi^{-1} (Y_i - g^{-1}(Z_1(x_i)^T \beta) - \gamma_i Z_2(x_i)^T \alpha) \right. \\
&\quad \left. | Y_i, x_i, \tilde{\beta}, \tilde{\alpha}, \tilde{\Psi} \right) \\
&= np \log(2\pi) + n \log |\Psi| + \sum_{i=1}^n e_i^T \Psi^{-1} e_i - 2e_i^T \Psi^{-1} (Z_2(x_i)^T \alpha) E \left(\gamma_i | Y_i, x_i, \tilde{\beta}, \tilde{\alpha}, \tilde{\Psi} \right) \\
&\quad + E \left(\gamma_i^2 | Y_i, x_i, \tilde{\beta}, \tilde{\alpha}, \tilde{\Psi} \right) (Z_2(x_i)^T \alpha)^T \Psi^{-1} (Z_2(x_i)^T \alpha)
\end{aligned}$$

mit $e_i = Y_i - g^{-1}(Z_1(x_i)^T \beta)$. Wegen

$$E \left(\gamma_i | Y_i, x_i, \tilde{\beta}, \tilde{\alpha}, \tilde{\Psi} \right) = m_i$$

und

$$\begin{aligned} E\left(\gamma_i^2 \mid Y_i, x_i, \tilde{\beta}, \tilde{\alpha}, \tilde{\Psi}\right) &= \left(E(\gamma_i \mid Y_i, x_i, \tilde{\beta}, \tilde{\alpha}, \tilde{\Psi})\right)^2 + \text{Var}\left(\gamma_i \mid Y_i, x_i, \tilde{\beta}, \tilde{\alpha}, \tilde{\Psi}\right) \\ &= m_i^2 + v_i \end{aligned}$$

ergibt sich

$$\begin{aligned} \tilde{l}(\alpha, \beta, \Psi) &= n(p+1) \log(2\pi) + \sum_{i=1}^n (m_i^2 + v_i) + n \log|\Psi| + \\ &\quad \sum_{i=1}^n e_i^T \Psi^{-1} e_i - 2m_i e_i^T \Psi^{-1} (Z_2(x_i)^T \alpha) + (v_i + m_i^2) (Z_2(x_i)^T \alpha)^T \Psi^{-1} (Z_2(x_i)^T \alpha). \end{aligned}$$

Mit $s_i = v_i^{\frac{1}{2}}$ folgt

$$\begin{aligned} \tilde{l}(\alpha, \beta, \Psi) &= n(p+1) \log(2\pi) + \sum_{i=1}^n (m_i^2 + v_i) + n \log|\Psi| + \\ &\quad \sum_{i=1}^n (e_i - m_i Z_2(x_i)^T \alpha)^T \Psi^{-1} (e_i - m_i Z_2(x_i)^T \alpha) + (s_i Z_2(x_i)^T \alpha)^T \Psi^{-1} (s_i Z_2(x_i)^T \alpha). \end{aligned}$$

Um die weitere Darstellung zu erleichtern, werden nun folgende Bezeichnungen eingeführt (ähnlich wie in Hoff und Niu (2012))

$$\begin{aligned} \theta &:= (\beta^T \ \alpha^T)^T \in \mathbb{R}^{q+w}, \\ \tilde{X}_i &:= \begin{pmatrix} Z_1(x_i)^T & 0_{p \times w} \end{pmatrix}, \\ \tilde{Z}_i^1 &:= \begin{pmatrix} 0_{p \times q} & m_i Z_2(x_i)^T \end{pmatrix} \text{ und} \\ \tilde{Z}_i^2 &:= \begin{pmatrix} 0_{p \times q} & s_i Z_2(x_i)^T \end{pmatrix}. \end{aligned} \tag{5.46}$$

Mit diesen Bezeichnungen erhält man

$$\begin{aligned} \tilde{l}(\theta, \Psi) &= n(p+1) \log(2\pi) + \sum_{i=1}^n (m_i^2 + v_i) + n \log|\Psi| + \\ &\quad \sum_{i=1}^n (Y_i - g^{-1}(\tilde{X}_i \theta) - \tilde{Z}_i^1 \theta)^T \Psi^{-1} (Y_i - g^{-1}(\tilde{X}_i \theta) - \tilde{Z}_i^1 \theta) \\ &\quad + (\tilde{Z}_i^2 \theta)^T \Psi^{-1} (\tilde{Z}_i^2 \theta). \end{aligned} \tag{5.47}$$

2. M-step

Im zweiten Schritt (M-step) wird die erwartete negative Log-Likelihoodfunktion (5.47) mithilfe des Newton-Algorithmus mit Schrittweitensuche minimiert. Dies geschieht zweischnittig, zunächst für den Parameter θ und anschließend für den Parameter Ψ , ähnlich wie bei Hoff und Niu (2012). Dazu werden der Gradient $\frac{\partial}{\partial \theta} \tilde{l}(\theta, \Psi)$, die Hesse-Matrix $\frac{\partial^2}{\partial \theta \partial \theta^T} \tilde{l}(\theta, \Psi)$ und die Ableitung $\frac{\partial}{\partial \Psi} \tilde{l}(\theta, \Psi)$ berechnet. Zunächst wird für ein festes Ψ ein $\hat{\theta}$ gesucht, für das gilt $\frac{\partial}{\partial \theta} \tilde{l}(\theta, \Psi) \Big|_{\theta=\hat{\theta}} = 0_q$ und anschließend wird ein $\hat{\Psi}$ ermittelt, welches $\frac{\partial}{\partial \Psi} \tilde{l}(\theta, \Psi) = 0_{p \times p}$ erfüllt. Es ist möglich, im Maximierung-Schritt lediglich einen Newton-Schritt mit anschließender Schrittweitensuche auszuführen, der zu einer Verringerung der erwarteten negativen Log-Likelihoodfunktion führt. Dieses Verfahren wird als ECM-Algorithmus bezeichnet und gehört zur Klasse der generalisierten EM-Algorithmen (McLachlan und Krishnan, 2008).

Der Gradient $\frac{\partial}{\partial \theta} \tilde{l}(\theta, \Psi)$ ergibt sich mithilfe der Ketten- und Produktregel zu

$$\frac{\partial}{\partial \theta} \tilde{l}(\theta, \Psi) = \sum_{i=1}^n \left\{ 2 \left(-(\tilde{X}_i)^T \begin{pmatrix} \frac{\partial g_1^{-1}}{\partial \eta_1} \Big|_{[\tilde{X}_i, \theta]_1} & 0 & \dots & 0 \\ \vdots & \ddots & & \vdots \\ 0 & \dots & 0 & \frac{\partial g_p^{-1}}{\partial \eta_p} \Big|_{[\tilde{X}_i, \theta]_p} \end{pmatrix} - \tilde{Z}_i^1{}^T \right) \cdot \Psi^{-1}(Y_i - g^{-1}(\tilde{X}_i \theta) - \tilde{Z}_i^1 \theta) + 2(\tilde{Z}_i^2)^T \Psi^{-1}(\tilde{Z}_i^2 \theta) \right\}. \quad (5.48)$$

Für $r \in \{1, \dots, p\}$ lautet der r -te Eintrag des Gradienten somit

$$\frac{\partial}{\partial [\theta]_r} \tilde{l}(\theta, \Psi) = \sum_{i=1}^n \left\{ 2 [(\tilde{Z}_i^2)^T \Psi^{-1} \tilde{Z}_i^2 \theta]_r + 2 \sum_{j_1=1}^p \sum_{j_2=1}^p (-[\tilde{X}_i^T]_{r, j_1} \frac{\partial g_{j_1}^{-1}}{\partial \eta_{j_1}} \Big|_{[\tilde{X}_i, \theta]_{j_1}} - [\tilde{Z}_i^1{}^T]_{r, j_1}) \cdot [\Psi^{-1}]_{j_1, j_2} ([Y_i]_{j_2} - g_{j_2}^{-1}(\tilde{X}_i \theta) - [\tilde{Z}_i^1 \theta]_{j_2}) \right\}. \quad (5.49)$$

Die einzelnen Einträge der Hesse-Matrix $\frac{\partial^2}{\partial \theta \partial \theta^T} \tilde{l}(\theta, \Psi)$ ergeben sich für $r, r' \in \{1, \dots, p\}$

mithilfe der Ketten- und Produktregel zu

$$\begin{aligned}
\frac{\partial^2}{\partial[\theta]_r \partial[\theta]_{r'}} \tilde{l}(\theta, \Psi) &= \sum_{i=1}^n \left\{ 2 \sum_{j_1=1}^p \sum_{j_2=1}^p \left((-[\tilde{X}_i^T]_{r,j_1} [\tilde{X}_i^T]_{r,j_2} \frac{\partial g_{j_1}^{-1}}{\partial \eta_{j_1}} \Big|_{[\tilde{X}_i\theta]_{j_1}}) \right. \right. \\
&\cdot [\Psi^{-1}]_{j_1,j_2} ([Y_i]_{j_2} - g_{j_2}^{-1}([\tilde{X}_i\theta]_{j_2}) - [\tilde{Z}_i^1\theta]_{j_2}) + 2(-[\tilde{X}_i^T]_{r,j_1} \frac{\partial g_{j_1}^{-1}}{\partial \eta_{j_1}} \Big|_{[\tilde{X}_i\theta]_{j_1}} - [\tilde{Z}_i^1]_{r,j_1}) \\
&\cdot [\Psi^{-1}]_{j_1,j_2} (-[\tilde{X}_i^T]_{r',j_2} \frac{\partial g_{j_2}^{-1}}{\partial \eta_{j_2}} \Big|_{[\tilde{X}_i\theta]_{j_2}} - [\tilde{Z}_i^1]_{r',j_2}) \left. \right) + 2[(\tilde{Z}_i^2)^T \Psi^{-1} \tilde{Z}_i^2]_{r,r'} \left. \right\} \quad (5.50)
\end{aligned}$$

Damit kann für ein festes Ψ eine Nullstelle von $\frac{\partial}{\partial \theta} \tilde{l}(\theta, \Psi)$ mithilfe des Newton-Algorithmus ermittelt werden.

Die Ableitung $\frac{\partial}{\partial \Psi} \tilde{l}(\theta, \Psi)$ berechnet sich mit den Regeln $\frac{\partial}{\partial A} \log(|\det(A)|) = (A^{-1})^T$ und $\frac{\partial}{\partial A} b^T A^{-1} b = -(A^{-1} + (A^{-1})^T) b$ mit $A \in \mathbb{R}^{p \times p}$, $b \in \mathbb{R}^p$ (Schmid und Trenkler, 2006) als

$$\begin{aligned}
\frac{\partial}{\partial \Psi} \tilde{l}(\theta, \Psi) &= n \Psi^{-1} - \sum_{i=1}^n \left\{ \Psi^{-1} (Y_i - g^{-1}(\tilde{X}_i\theta)) \right. \\
&\quad \left. - \tilde{Z}_i^1\theta (Y_i - g^{-1}(\tilde{X}_i\theta) - \tilde{Z}_i^1\theta)^T \Psi^{-1} + \Psi^{-1} (\tilde{Z}_i^2\theta) (\tilde{Z}_i^2\theta)^T \Psi^{-1} \right\} \quad (5.51)
\end{aligned}$$

Setzt man die Gleichung (5.51) gleich Null, so folgt als Lösung

$$\hat{\Psi} = \frac{1}{n} \sum_{i=1}^n \left\{ (Y_i - g^{-1}(\tilde{X}_i\theta) - \tilde{Z}_i^1\theta) (Y_i - g^{-1}(\tilde{X}_i\theta) - \tilde{Z}_i^1\theta)^T + (\tilde{Z}_i^2\theta) (\tilde{Z}_i^2\theta)^T \right\} \quad (5.52)$$

Abschließend erfolgt eine Zusammenfassung der Schritte zur Schätzung der Koeffizientenvektoren β und α für das Kovarianzregressionsmodell in folgendem vollständigem Algorithmus.

Algorithmus 5.6 (Anpassung des Kovarianzregressionsmodells)

1. *Bestimme Startvektoren $\beta^{(0)}$ und $\alpha^{(0)}$. Lege Abbruchkonstante $\epsilon > 0$ und Konstante $\delta > 0$ zur Schrittweitensteuerung fest. Setze $\theta^{(0)} = (\beta^{(0)}, \alpha^{(0)})$ und $s = 0$.*

2. **Schritt ($s+1$):**

E-Step: Berechne für $i = 1, \dots, n$

(a) die Ausdrücke m_i, v_i gemäß (5.45) sowie

(b) die Matrizen $\tilde{X}_i, \tilde{Z}_i^1$ und \tilde{Z}_i^2 gemäß (5.46).

M-Step: Berechne einen Newton-Schritt

$$d^{(s+1)} = \left(\frac{\partial^2}{\partial \theta \partial \theta^T} \tilde{l}(\theta, \Psi^{(s)}) \right)^{-1} \Big|_{\theta=\theta^{(s)}} \frac{\partial}{\partial \theta} \tilde{l}(\theta, \Psi^{(s)}) \Big|_{\theta=\theta^{(s)}}$$

gemäß (A.1) und suche Schrittweite $t^{(s+1)}$ mit $\tilde{l}(\theta^{(s)} + t^{(s+1)}d^{(s+1)}, \Psi^{(s)}) < \tilde{l}(\theta^{(s)}, \Psi^{(s)})$.

Setze $\theta^{(s+1)} = \theta^{(s)} + t^{(s+1)}d^{(s+1)}$. Berechne anschließend $\Psi^{(s+1)}$ gemäß (5.52) für festes $\theta = \theta^{(s+1)}$.

3. Wird ein bestimmtes Abbruchkriterium erreicht, wie etwa $|l(\beta^{(s+1)}, \alpha^{(s+1)}, \Psi^{(s+1)}) - l(\beta^{(s)}, \alpha^{(s)}, \Psi^{(s)})| < \epsilon$, so gehe zu 4. Andernfalls setze $s = s + 1$ und gehe zu 2.
4. Die geschätzten Parameter im Kovarianzregressionsmodell lauten $\hat{\beta} = \beta^{(s+1)}$, $\hat{\alpha} = \alpha^{(s+1)}$ und $\hat{\Psi} = \Psi^{(s+1)}$.

Der M-step besteht wie bereits erwähnt aus zwei Schritten. Für gegebenes $\Psi^{(s)}$ wird ein $\theta^{(s+1)}$ berechnet und anschließend ein $\Psi^{(s+1)}$ für festes $\theta^{(s+1)}$. Der EM-Algorithmus wird typischerweise fortgeführt, bis die Differenz der aufeinanderfolgenden Funktionswerte der Likelihoodfunktion eine vorgegebene Grenze unterschreitet (McLachlan und Krishnan, 2008; Hoff und Niu, 2012). Für den Fall einer Linkfunktion der Art $g : \eta \rightarrow \eta$ kann die Minimierung der erwarteten negativen Log-Likelihoodfunktion (5.47) analytisch direkt bestimmt werden. Damit wird der M-step im Schritt $(s + 1)$ im Algorithmus 5.6 zu

$$\theta^{(s+1)} = \left(\sum_{i=1}^n (\tilde{X}_i - \tilde{Z}_i^1)^T \Psi^{-1} (\tilde{X}_i - \tilde{Z}_i^1) + \tilde{Z}_i^{2T} \Psi^{-1} \tilde{Z}_i^2 \right)^{-1} \left(\sum_{i=1}^n \Psi^{-1} (\tilde{X}_i - \tilde{Z}_i^1) Y_i \right)$$

$$\Psi^{s+1} = \frac{1}{n} \sum_{i=1}^n \left((Y_i - (\tilde{X}_i - \tilde{Z}_i^1)\theta)(Y_i - (\tilde{X}_i - \tilde{Z}_i^1)\theta)^T + (\tilde{Z}_i^2\theta)(\tilde{Z}_i^2\theta)^T \right).$$

5.4. Heuristischer Ansatz

In diesem Abschnitt wird zum Vergleich der bereits eingeführten Modelle eine Heuristik hinzugenommen, die aus folgenden Schritten besteht.

1. Bestimme generalisierte lineare Modelle für jede einzelne Zielgröße $[Y]_j, j, \dots, p$:
 - Normalverteilungsannahme für $([Y]_j|x)$,
 - Strukturkomponenten $\eta_j = f_j(x)^T \beta_j$ und
 - Linkfunktion $\eta_j = g^{-1}(\mu_j)$.

Bezeichne mit $\hat{\beta}_j$ den geschätzten Koeffizientenvektor β_j und mit $\hat{\mu}_{mat} \in \mathbb{R}^{n \times p}$ die Matrix der vorhergesagten Werte, wobei die i -te Zeile den vorhergesagten Erwartungswertvektor für den Kovariablenvektor x_i enthält.

2. Die Kovarianzmatrix wird anschließend nichtparametrisch wie in Kapitel 5.1 geschätzt über

$$\hat{\Sigma}(x) = \left[\sum_{i=1}^n K_h(x_i - x) \{Y_i - [\hat{\mu}_{mat}]_{i,\cdot}\} \{Y_i - [\hat{\mu}_{mat}]_{i,\cdot}\}^T \right] \left\{ \sum_{i=1}^n K_h(x_i - x) \right\}. \quad (5.53)$$

mit einer Kernfunktion $K_h : \mathbb{R}^k \rightarrow \mathbb{R}$ und einer Bandbreite $h > 0$. Dabei wird die Kernfunktion (5.5) verwendet.

3. Die Bestimmung einer geeigneten Bandbreite erfolgt durch Minimierung von

$$CV_{\Sigma}(h) = \frac{1}{n} \sum_{i=1}^n \left\{ (Y_i - [\hat{\mu}_{mat}]_{-i})^T \hat{\Sigma}_{-i}(x_i)^{-1} (Y_i - [\hat{\mu}_{mat}]_{-i}) - \log(|\hat{\Sigma}_{-i}(x_i)^{-1}|) \right\}.$$

4. Das Vorhersagemodell lautet nun

$$\begin{aligned} \hat{\mu}_j(x) &= g^{-1}(f_j(x)^T \hat{\beta}_j) \text{ für } j = 1, \dots, p \\ \hat{\Sigma}(x) &= \left[\sum_{i=1}^n K_h(x_i - x) \left\{ Y_i - [\hat{\mu}_{mat}]_{i,\cdot} \right\} \left\{ Y_i - [\hat{\mu}_{mat}]_{i,\cdot} \right\}^T \right] \left\{ \sum_{i=1}^n K_h(x_i - x) \right\}. \end{aligned} \quad (5.54)$$

Dieser heuristische Ansatz kombiniert die parametrische und nichtparametrische Modellierung und wird in der Simulationsstudie im nächsten Kapitel zum Vergleich aufgenommen. Insbesondere bei Vorliegen von Kovariablen und einer Struktur der Form (5.10) wird hier eine bessere Schätzung als bei dem Nadaraya-Watson-Kernschätzer aus Kapitel 5.1 erwartet.

5.5. Simulationsstudie zum Vergleich der Methoden

Die in Kapitel 5 vorgestellten Ansätze zur Modellierung und Schätzung von Kovarianzmatrizen sollen nachfolgend in einer Simulationsstudie bezüglich ihrer Vorhersagegenauigkeit verglichen werden. Zusätzlich zu den bereits beschriebenen Modellen wird bei vorliegenden Wiederholungen ein einfaches Vorgehen hinzugezogen, welches lediglich den empirischen Erwartungswert

$$\mu(x_i) = \frac{1}{n_i} \sum_{j=1}^{n_i} y_j^i$$

und die empirische Kovarianzmatrix

$$\Sigma(x_i) = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (y_j^i - \mu(x_i))(y_j^i - \mu(x_i))^T$$

pro Designvektor x_i berechnet. Im folgenden werden die zu vergleichenden Modelle wie folgt bezeichnet:

M1: Kovarianzregressionsmodell aus Kapitel 5.3

M2: Matrix-Logarithmus-Modell aus Kapitel 5.2

M3: Nichtparametrisches Modell aus Kapitel 5.1

M4: Heuristisches Modell aus Kapitel 5.4

M5: Einfaches Vorgehen mit empirischem Erwartungswert und empirischer Kovarianzmatrix pro Designvektor x_i

Die Schätzverfahren sollen anhand verschiedener Datensituationen verglichen werden. Motiviert werden diese durch die Fragestellung der zweistufigen Modellierung $X \rightarrow Y \rightarrow Z$ des thermischen Spritzprozesses (siehe Kapitel 2). Die Tatsache, dass auf der $(X \rightarrow Y)$ -Ebene ein klassischer Versuchsplan festgelegt werden kann, wird durch Verwendung eines zentral zusammengesetzten Versuchsplans beachtet. Diese Wahl wird mit **Design 1** bezeichnet. Auf der $(Y \rightarrow Z)$ -Ebene sind die Y -Werte eher zufälliger Natur. Daher werden Werte von Kovariablen zufällig generiert und mit **Design 2** bezeichnet. Es werden Fälle mit 2, 3 oder 4 Kovariablen und mit 2 oder 3 Zielgrößen untersucht. Die wahre Erwartungswertfunktion wird mit dem linearen Prädiktor $\eta = Z_1(x)^T \beta$ entweder mittels eines Identitätslinks über

$$\mu^{true}(x) = Z_1(x)^T \beta$$

oder mithilfe eines Logarithmuslinks über

$$\mu^{true}(x) = g^{-1}(Z_1(x)^T \beta)$$

dargestellt, wobei $g^{-1} : \eta \in \mathbb{R}^p \rightarrow (\exp([\eta]_1), \dots, \exp([\eta]_p)) \in \mathbb{R}^p$ ist. Die Matrix $Z_1(x)$ ist vom Typ (5.9), wobei die einzelnen Vektoren $f_i(x) = (1, [x]_1, \dots, [x]_k)$ Haupteffekte enthalten. Des Weiteren werden drei unterschiedliche Typen einer Kovarianzmatrix untersucht, wie folgt:

Typ 1: Kovarianzmatrizen vom Typ wie in Modell (5.12) mit

$$\Sigma_{typ1}^{true}(x_i) = \exp(A(x_i)) \quad (= \exp(\sum_{j=1}^g [\alpha]_j U_j(x_i))).$$

Die einzelnen Einträge in $A(x_i)$ seien lineare Modelle mit Haupteffekten, also $[A(x_i)]_{ij} = [\alpha]_1 + [\alpha]_2[x]_1 + \dots + [\alpha]_{k+1}[x]_k$.

Typ 2: Kovarianzmatrizen vom Typ (5.31) mit

$$\Sigma_{typ2}^{true}(x_i) = \Psi + (Z_2(x_i)^T \alpha)(Z_2(x_i)^T \alpha)^T.$$

Die Matrizen $Z_2(x_i)$ sind wie in (5.30) gewählt, wobei die einzelnen Vektoren $\tilde{f}_i(x) = (1, [x]_1, \dots, [x]_k)$ nur Haupteffekte enthalten.

Typ 3: Kovarianzmatrizen vom Cholesky-Typ

$$\Sigma_{typ3}^{true}(x_i) = C(x_i)C(x_i)^T$$

mit unteren Dreiecksmatrizen $C(x_i)$ mit Einträgen

$$[C(x_i)]_{lj} = \begin{cases} 0 & , l < j \\ [\alpha_{lj}]_1 + [\alpha_{lj}]_2[x]_1 + \dots + [\alpha_{lj}]_{k+1}[x]_k & , l \leq j. \end{cases}$$

Die festgelegten Koeffizientenvektoren $\beta, \alpha, \alpha_{ij}$ ($i \leq j = 1, \dots, p$) sowie die gewählten Matrizen $U_j(x_i)$ sind im Anhang A.3 zu finden. Seien $\mu^{true}(x)$ sowie $\Sigma^{true}(x)$ die jeweiligen festgelegten Momente. Die Güte der Schätzmethode wird für die geschätzten Momente $\hat{\mu}(x)$ und $\hat{\Sigma}(x)$ anhand der Distanzen zwischen wahren und geschätzten Erwartungswertvektoren

$$\Delta_\mu = \|\hat{\mu}(x) - \mu^{true}(x)\|_2$$

sowie wahren und geschätzten Kovarianzmatrizen

$$\Delta_\Sigma = \|\hat{\Sigma}(x) - \Sigma^{true}(x)\|_F$$

für $i = 1, \dots, n$ untersucht. Hier steht $\|\cdot\|_2$ für die euklidische Norm und $\|\cdot\|_F$ für die Frobeniusnorm.

Der Ablauf der Simulationsstudie für $n_x = 2, 3, 4$, $n_y = 2, 3$, jeden Kovarianztyp 1, 2, 3 und den Identitätslink sowie den Logarithmuslink kann nun wie folgt zusammengefasst werden.

1. Kovariablen:

Zur Erzeugung von Datensätzen mit Kovariablen wird wie folgt vorgegangen.

- (a) Es wird Design 1 als orthogonal geblockter zentral zusammengesetzter Versuchsplan mit 8 Zentralpunkten für n_x Kovariablen erzeugt und anschließend 0, 1, 2, 3, 4 mal wiederholt. Bezeichne mit $\mathcal{D}_{n_x, r}^1$ das r -mal wiederholte Design 1 mit n_x Kovariablen. Bezeichne die Anzahl der Einstellungen von $\mathcal{D}_{n_x, r}^1$ mit n .

- (b) Es wird Design 2 mit $n = 20, 30, 40, \dots, 100$ Einstellungen erzeugt. Bezeichne mit $\mathcal{D}_{n_x, n}^2$ das Design 2 mit n_x Kovariablen und Stichprobengröße n .

2. Momente der Zielgrößen:

Ausgehend von einem generierten Datensatz $\mathcal{D}_{n_x, r}^1$ oder $\mathcal{D}_{n_x, n}^2$ werden die wahren Momente basierend auf den Kovarianztypen 1, 2 oder 3 und der gewählten Linkfunktion erzeugt, indem die Koeffizientenvektoren β sowie die Koeffizientenvektoren für den jeweiligen Kovarianztyp festgelegt werden und anschließend die wahren Erwartungswertvektoren $\mu^{true}(x_i)$ und Kovarianzmatrizen $\Sigma^{true}(x_i)$ berechnet werden.

3. Zielgrößenwerteschätzung der Modelle:

Für $j = 1, \dots, 100$

1. wird eine zufällige Stichprobe der Größe n aus der multivariaten Normalverteilung auf Basis der wahren Erwartungswertvektoren $\mu^{true}(x_i)$ und Kovarianzmatrizen $\Sigma^{true}(x_i)$ ($i = 1, \dots, n$) gezogen.
2. werden die Erwartungswertvektoren und die Kovarianzmatrizen für die Modelle M1 bis M5 geschätzt.
3. werden für jedes der Modelle M1 bis M5 die Distanzen Δ_μ und Δ_Σ berechnet.

4. Ergebnisdarstellung:

Erstellung von Boxplots für die Distanzen Δ_μ und Δ_Σ .

Im folgenden werden die Ergebnisse dargestellt.

Ergebnisse für Kovarianztyp 1

In Abbildung 5.1 sind die Simulationsergebnisse für den Kovarianztyp 1, 2 Kovariablen, 2 Zielgrößen und den Identitätslink unter Verwendung des Design 1 dargestellt. Die linke Abbildung bildet die Werte Δ_μ auf der vertikalen Achse gegen die Wiederholungen

des Design 1 (horizontale Achse) ab. In der rechten Abbildung werden die Werte Δ_Σ dargestellt. Die Distanzen Δ_μ und Δ_Σ werden für jedes der Modelle M1 bis M5 farblich getrennt in Form von Boxplots abgebildet, wobei Werte außerhalb des 1,5-fachen Interquartilsabstands in dieser Abbildung nicht dargestellt werden. Bei perfekter Anpassung sind die Distanzen Δ_μ und Δ_Σ gleich Null. Somit sollten die Boxplots möglichst nahe bei Null und möglichst klein sein. Des Weiteren dient das Vorgehen M5 als einfacher Vergleich und es wird erwartet, dass die restlichen Modelle M1 bis M4 mindestens genauso gut im Hinblick auf die Schätzung des Erwartungswertvektors und der Kovarianzmatrix sind wie M5.

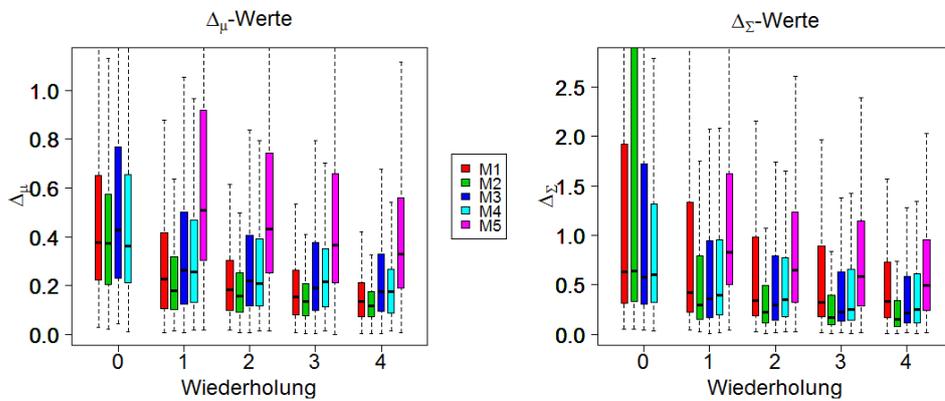


Abbildung 5.1: Boxplots der Distanzen Δ_μ und Δ_Σ für den Kovarianztyp 1, 2 Kovariablen, 2 Zielgrößen, Identitätslink und Design 1

In Abbildung 5.1 ist zunächst zu sehen, dass die Boxen für M1 bis M4 deutlich unterhalb der Box für M5 liegen. Insgesamt ist die Schätzung basierend auf Modell M2 am besten. Dies ist nicht überraschend, da der Kovarianztyp 1 dem Kovarianzmodell aus M2 entspricht. In der rechten Abbildung ist im Fall ohne Wiederholungen zu erkennen, dass die Box für M2 sehr breit ist, was auf mögliche Konvergenzprobleme hindeuten könnte. Dies kann unter Umständen durch eine bessere Wahl der Startwerte gelöst werden. Außerdem ist zu beobachten, dass die Schätzung für wachsende Anzahl an Wiederholungen des Versuchsplans besser wird. Die Boxen nähern sich der Null und werden zudem schmaler. Bei Hinzunahme von weiteren Kovariablen und einer weiteren Zielgröße ergibt

sich ein Bild wie zuvor, wobei die Schätzung insbesondere bezüglich Δ_μ basierend auf Modell M2 deutlich besser wird, wie in Abbildung 5.2 zu sehen ist. Hier ist zudem zu

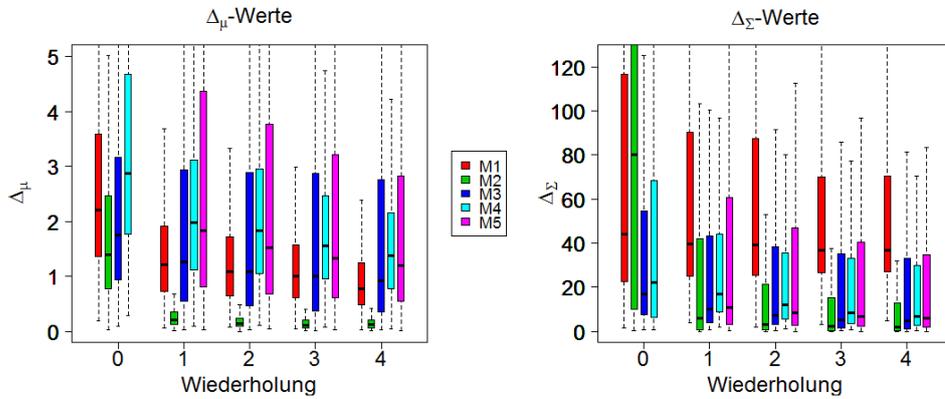


Abbildung 5.2: Boxplots der Distanzen Δ_μ und Δ_Σ für den Kovarianztyp 1, 4 Kovariablen, 3 Zielgrößen, Identitätslink und Design 1

beobachten, dass das Modell M1 durchgehend schlechter als M5 ist. Der Kovarianztyp 1 ist eine hochgradig nichtlineare Transformation und das Modell M1 unterstellt ein relativ einfaches Modell für die Kovarianzmatrix. Das könnte die schlechtere Schätzung der Kovarianzmatrix erklären. Der Erwartungswertvektor wird dennoch durch das Modell M1 und M2 am besten geschätzt. Die Boxen in der rechten Grafik der Abbildung 5.2 werden mit wachsenden Wiederholungen schmaler. Beispielhaft stellt Abbildung 5.3 die Simulationsergebnisse für den Kovarianztyp 1, 4 Kovariablen, 2 Zielgrößen und den Logarithmuslink unter Verwendung von Design 1 dar. Hier lässt sich wieder die Überlegenheit des Modells M2 wie erwartet erkennen. Bei den Versuchen ohne Wiederholung schneidet das Modell M2 schlechter ab als M1 und M4, allerdings ist die Schätzung ab Wiederholung 1 am besten. Lediglich die Modelle M2 und M4 sind durchgehend besser als das einfache Vorgehen M5.

Als nächstes wird der Einfluss der Stichprobengröße bei zufällig generierten Kovariablenwerten auf die Distanzen Δ_μ und Δ_Σ untersucht. In Abbildung 5.4 sind die Ergebnisse für 2 Kovariablen und 2 Zielgrößen unter Verwendung des Identitätslinks beispielhaft dargestellt. Dabei wurde Design 2 verwendet und die Distanzen Δ_μ und Δ_Σ

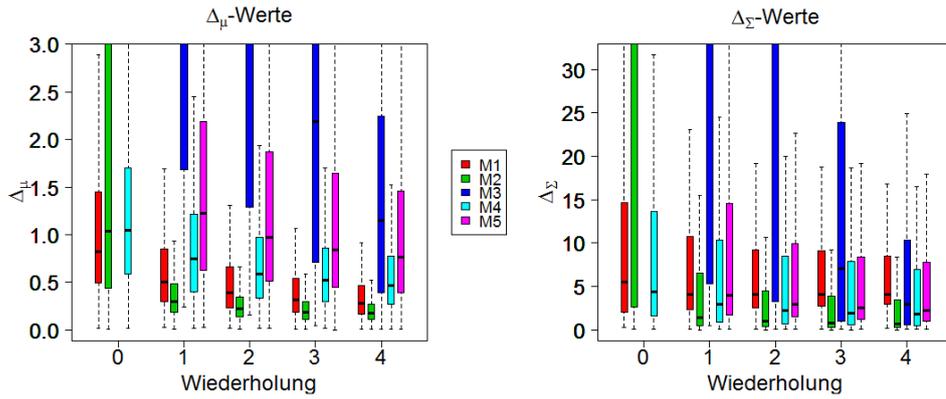


Abbildung 5.3: Boxplots der Distanzen Δ_μ und Δ_Σ für den Kovarianztyp 1, 4 Kovariablen, 2 Zielgrößen, Logarithmuslink und Design 1

für Stichprobengrößen 20, 30, 40, ..., 100 berechnet. Die Ergebnisse für das Modell M2

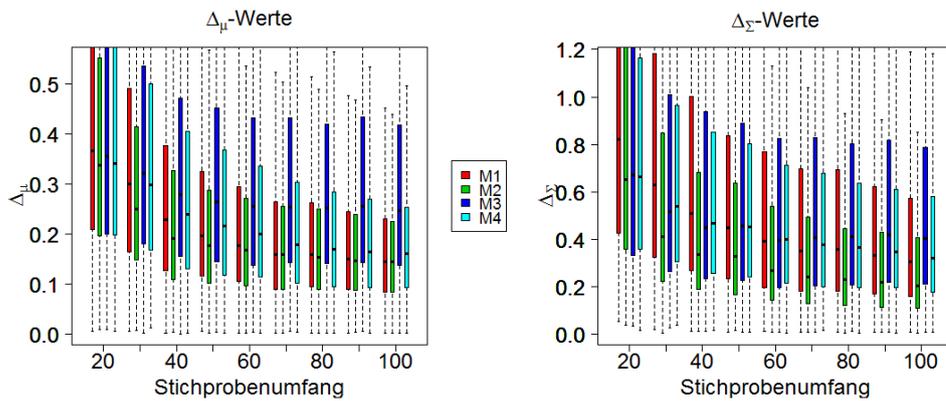


Abbildung 5.4: Boxplots der Distanzen Δ_μ und Δ_Σ für den Kovarianztyp 1, 2 Kovariablen, 2 Zielgrößen, Identitätslink und Design 2

werden schnell mit wachsender Stichprobengröße besser und sind den restlichen Modellen bereits ab Stichprobengröße 30 überlegen. Die schlechtesten Ergebnisse liefert das Modell M3, wobei M1 und M4 zu ähnlichen Werten bezüglich Δ_Σ führen. Mit wachsender Stichprobengröße werden auch die Schätzungen basierend auf M1 und M4 deutlich besser. Die Hinzunahme weiterer Kovariablen und einer weiteren Zielgröße führt zu ähnlichen Ergebnissen, wie in Abbildung 5.5 zu sehen ist. Hier werden die Schätzungen basierend auf M2 mit wachsender Stichprobengröße deutlich besser, auch wenn die

Schätzung für die Stichprobengrößen 20, 30 und 40 am schlechtesten ist.

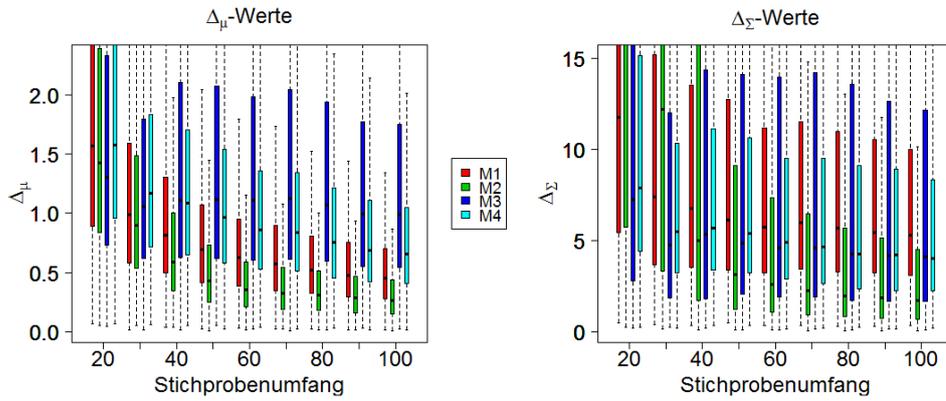


Abbildung 5.5: Boxplots der Distanzen Δ_μ und Δ_Σ für den Kovarianztyp 1, 4 Kovariablen, 3 Zielgrößen, Identitätslink und Design 2

In Abbildung 5.6 sind die Ergebnisse bei Verwendung des Logarithmuslinks beispielhaft für 4 Kovariablen und 2 Zielgrößen dargestellt. Die Schätzung basierend auf M3 versagt hier und liegt deutlich hinter den anderen Modellen zurück. Das heuristische Modell M4 liefert hier die zweitschlechtesten Resultate. Es ist zu erkennen, dass lediglich die Schätzungen der Kovarianzmatrix basierend auf dem Modell M2 mit wachsendem Stichprobenumfang besser werden. Ab Stichprobenumfang 40 sind die Schätzungen sowohl des Erwartungswertvektors als auch der Kovarianzmatrix basierend auf dem Modell M2 am besten und mit wachsendem Stichprobenumfang werden die Abstände zu den übrigen Modellen größer.

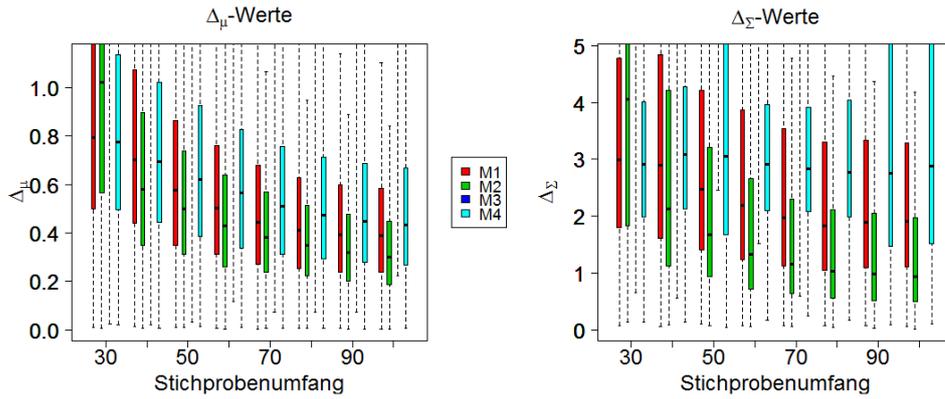


Abbildung 5.6: Boxplots der Distanzen Δ_μ und Δ_Σ für den Kovarianztyp 1, 4 Kovariablen, 2 Zielgrößen, Logarithmuslink und Design 2

Ergebnisse für Kovarianztyp 2

In Abbildung 5.7 sind die Simulationsergebnisse für den Kovarianztyp 2, 2 Kovariablen, 2 Zielgrößen und den Identitätslink unter Verwendung von Design 1 zu sehen. Bei den

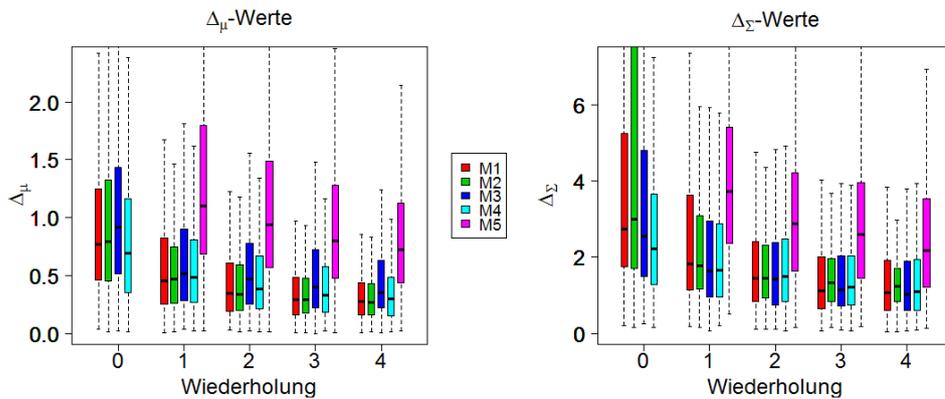


Abbildung 5.7: Boxplots der Distanzen Δ_μ und Δ_Σ für den Kovarianztyp 2, 2 Kovariablen, 2 Zielgrößen, Identitätslink und Design 1

Versuchen ohne Wiederholungen ist zu erkennen, dass das Modell M2 zu den schlechtesten Schätzungen führt. Die übrigen Modelle weisen eine ähnliche Performanz auf und sind in jedem Fall besser als das einfache Vorgehen M5. Mit wachsenden Wiederholungen werden die Schätzungen aller Modelle besser. Durch Hinzunahme weiterer Kovariablen

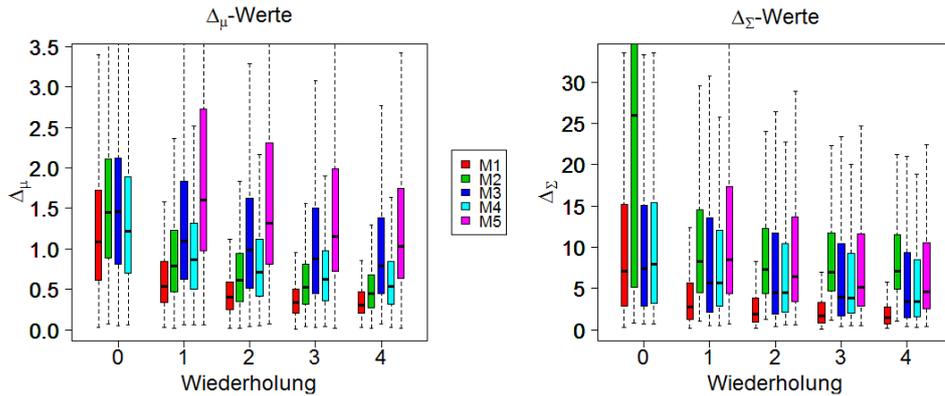


Abbildung 5.8: Boxplots der Distanzen Δ_μ und Δ_Σ für den Kovarianztyp 2, 4 Kovariablen, 3 Zielgrößen, Identitätslink und Design 1

und einer weiteren Zielgröße wird wie erwartet die Überlegenheit des Modells M1 deutlicher, wie in Abbildung 5.8 zu erkennen ist. Die Distanzen Δ_μ und Δ_Σ sind für das Modell M1 deutlich geringer und stabiler als für die übrigen Modelle. Bezüglich der Kovarianzmatrix ist die Schätzung basierend auf dem Modell M4 etwas besser als für die Modelle M2, M3 und M5. Der Erwartungswertvektor wird am besten durch das Modell M1 geschätzt. Die Modelle M1, M3 und M4 schätzen hier durchgehend besser als das Modell M5. Bei Verwendung des Logarithmuslinks ist, beispielhaft dargestellt in Abbildung 5.9, zu sehen, dass das Modell M1 die beste Performanz liefert und das nichtparametrische Modell M3 zu den schlechtesten Schätzungen führt. Lediglich die Modelle M1 und M4 führen hier zu Schätzungen, die sowohl bezüglich Δ_μ als auch Δ_Σ besser sind als für das Modell M5.

Als nächstes wird der Einfluss der Stichprobengröße ohne Wiederholungen untersucht. In Abbildung 5.10 sind die Simulationsergebnisse für den Kovarianztyp 2, 2 Kovariablen, 2 Zielgrößen und den Identitätslink unter Verwendung von Design 2 für unterschiedliche Stichprobengrößen (x-Achse) zu sehen. Es ist zu erkennen, dass alle betrachteten Modelle bezüglich ihrer Performanz mit wachsendem Stichprobenumfang besser werden. Bezüglich der Distanzen Δ_μ verhalten sich die Modelle M1, M2 und M4 ähnlich und bezüglich Δ_Σ liefert das Modell M2 ab Stichprobenumfang 40 die besten Ergebnisse. Bei

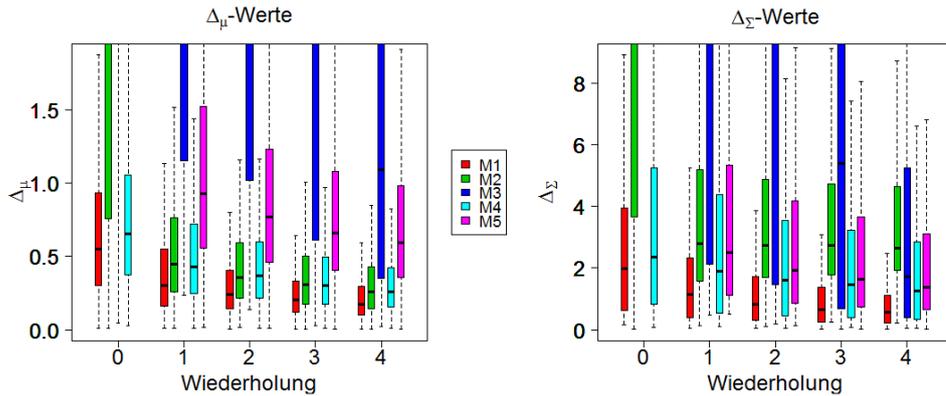


Abbildung 5.9: Boxplots der Distanzen Δ_μ und Δ_Σ für den Kovarianztyp 2, 4 Kovariablen, 2 Zielgrößen, Logarithmuslink und Design 1

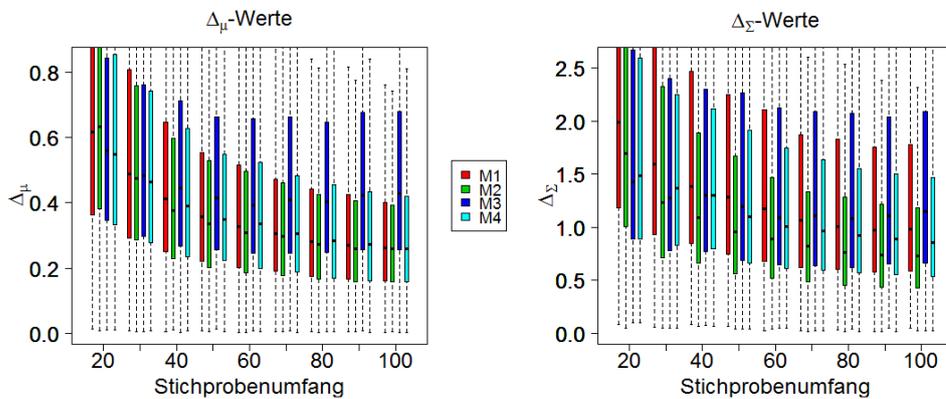


Abbildung 5.10: Boxplots der Distanzen Δ_μ und Δ_Σ für den Kovarianztyp 2, 2 Kovariablen, 2 Zielgrößen, Identitätslink und Design 2

Stichprobenumfang 20 und 30 ist das Modell M4 etwas besser als die übrigen.

Abbildung 5.11 zeigt die Ergebnisse für den Fall von 4 Kovariablen und 3 Zielgrößen. Hier setzt sich ab Stichprobengröße 50 das Modell M1 deutlich durch. Mit wachsendem Stichprobenumfang wird lediglich die Schätzung basierend auf den Modellen M1, M2 und M4 besser. Das Modell M3 führt zu den schlechtesten Resultaten.

Abbildung 5.12 zeigt die Ergebnisse bei Verwendung des Logarithmuslinks beispielhaft für 4 Kovariablen und 2 Zielgrößen. Wie zuvor beobachtet, liegt die Schätzung basierend

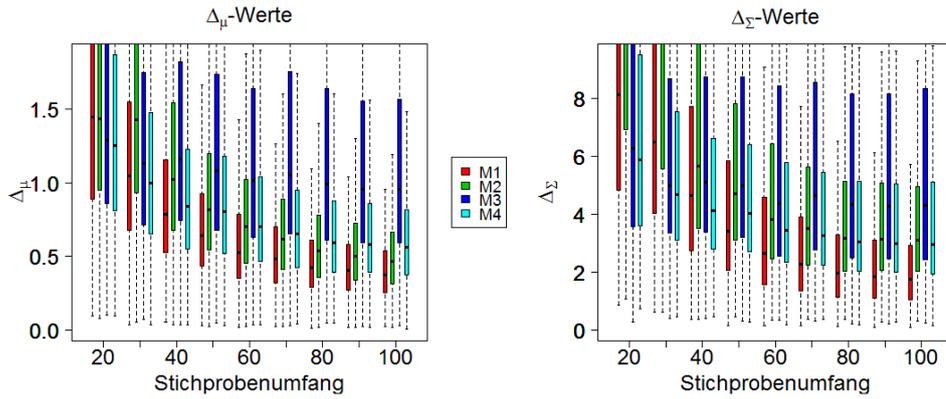


Abbildung 5.11: Boxplots der Distanzen Δ_μ und Δ_Σ für den Kovarianztyp 2, 4 Kovariablen, 3 Zielgrößen, Identitätslink und Design 2

auf dem Modell M3 deutlich hinter den übrigen zurück. Die Schätzung des Erwartungswertvektors wird für die Modelle M1 und M2 mit wachsender Stichprobengröße besser und ab Stichprobengrößen 60 liefert das Modell M1 die besten Resultate sowohl für den Erwartungswertvektor als auch für die Kovarianzmatrix. Die Modelle M1 und M2 liefern mit wachsendem Stichprobenumfang immer bessere Schätzungen der Kovarianzmatrix, wobei sich die Überlegenheit des Modells M1 deutlich ab Stichprobenumfang 60 zeigt.

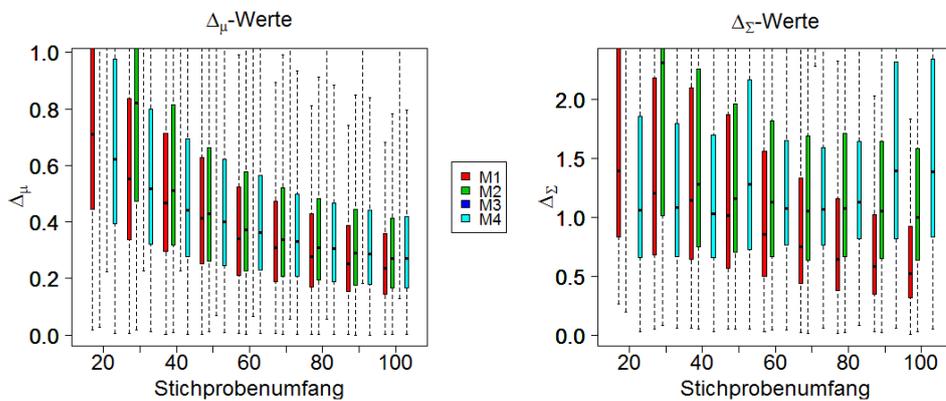


Abbildung 5.12: Boxplots der Distanzen Δ_μ und Δ_Σ für den Kovarianztyp 2, 4 Kovariablen, 2 Zielgrößen, Logarithmuslink und Design 2

Ergebnisse für Kovarianztyp 3

Abbildung 5.13 stellt die Simulationsergebnisse für den Kovarianztyp 3, 2 Kovariablen, 2 Zielgrößen und des Identitätslinks unter Verwendung von Design 1 dar. Wie bereits zuvor beobachtet, weist das Modell M2 die schlechtesten Schätzungen für die Versuche ohne Wiederholung bezüglich der Werte Δ_Σ auf. Für wachsende Wiederholungen werden die Ergebnisse für die Modelle M1 bis M4 besser und verhalten sich ähnlich. In jedem Fall wird das einfache Vorgehen M5 übertroffen. In Abbildung 5.14 sind die Ergebnisse

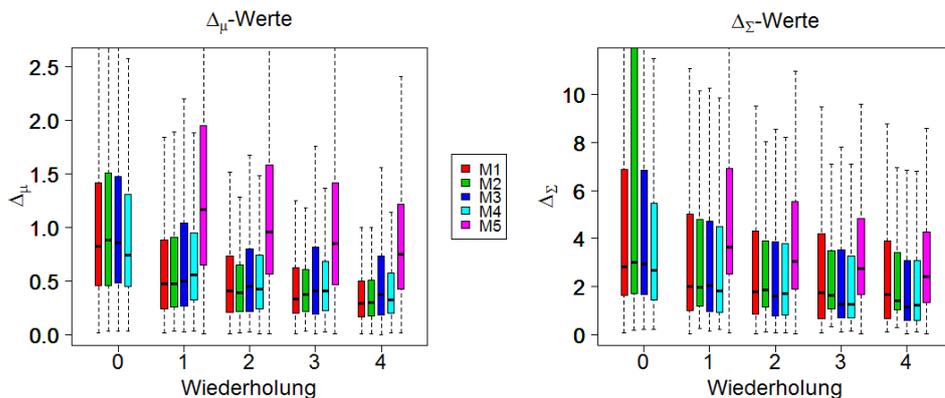


Abbildung 5.13: Boxplots der Distanzen Δ_μ und Δ_Σ für den Kovarianztyp 3, 2 Kovariablen, 2 Zielgrößen, Identitätslink und Design 1

für 4 Kovariablen und 3 Zielgrößen dargestellt. Hier wird deutlich, dass das Modell M1 zu etwas besseren und stabileren Schätzungen der Kovarianzmatrix führt. Im Median sind die Distanzen Δ_Σ für die Modelle M1 und M4 in etwa gleich, allerdings ist die Box für M1 deutlich kleiner als die Box für M4. Bezüglich der Distanzen Δ_μ liefert das Modell M1 etwas bessere Schätzungen als die übrigen Modelle. In jedem Fall führt das Modell M5 zu den schlechtesten Schätzungen. Bei Verwendung des Logarithmuslinks ist in Abbildung 5.15 zu beobachten, dass das nichtparametrische Modell M3 deutlich hinter allen anderen zurückliegt. Die Modelle M1 und M2 führen zu den besten Vorhersagen bezüglich Δ_μ . Im Median sind die Distanzen Δ_μ für die Modelle M2 und M4 ab Wiederholung 1 ähnlich. Bezogen auf die Werte Δ_Σ führen die Modelle M1 und M4 zu besseren Schätzungen als das einfache Vorgehen M5, wobei das Modell M1 am besten abschneidet. Des Weiteren

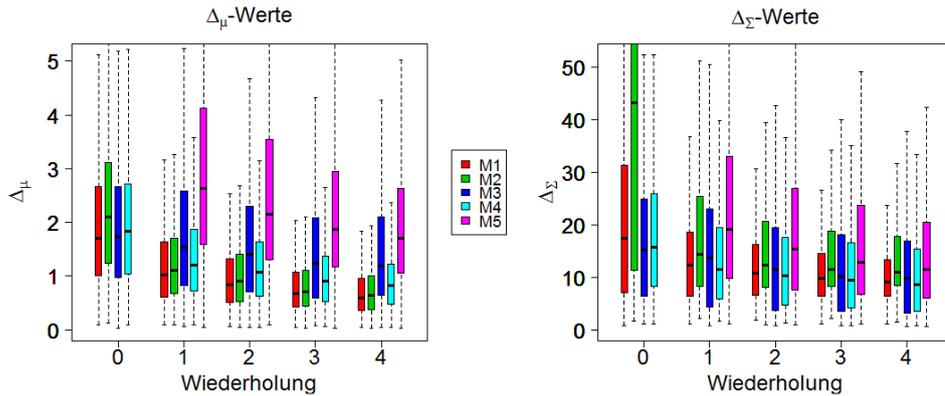


Abbildung 5.14: Boxplots der Distanzen Δ_μ und Δ_Σ für den Kovarianztyp 3, 4 Kovariablen, 3 Zielgrößen, Identitätslink und Design 1

werden die Boxen für M1 und M2 mit wachsenden Wiederholungen kleiner und damit die Schätzungen stabiler.

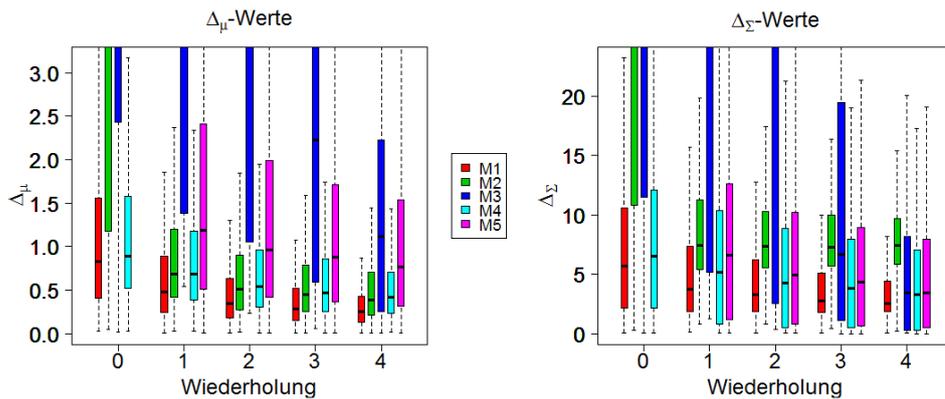


Abbildung 5.15: Boxplots der Distanzen Δ_μ und Δ_Σ für den Kovarianztyp 3, 4 Kovariablen, 2 Zielgrößen, Logarithmuslink und Design 1

Jetzt wird der Einfluss der Stichprobengröße ohne Wiederholungen untersucht. Abbildung 5.16 zeigt die Simulationsergebnisse für den Kovarianztyp 3, 2 Kovariablen, 2 Zielgrößen und den Identitätslink unter Verwendung von Design 2 für unterschiedliche Stichprobengrößen (x-Achse). Die Modelle M1, M2 und M4 werden bezüglich der Werte Δ_μ und Δ_Σ mit wachsendem Stichprobenumfang besser. Für die Werte Δ_Σ führt

das Modell M2 ab der Stichprobengröße 40 zu den besten Schätzungen. Für den Fall

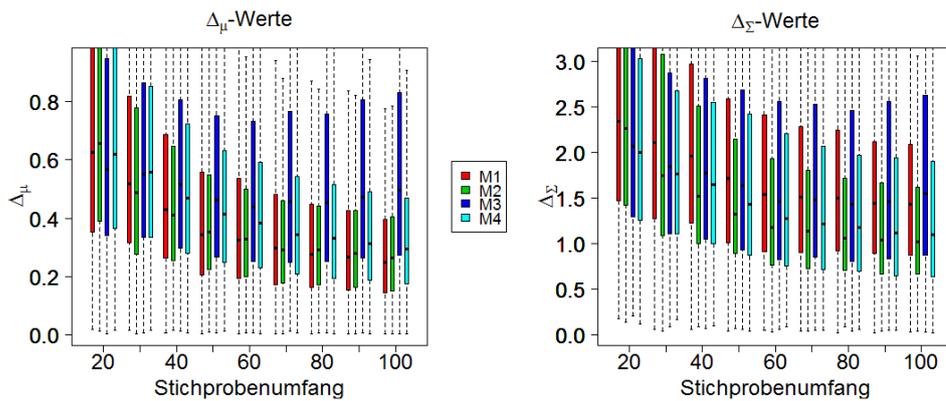


Abbildung 5.16: Boxplots der Distanzen Δ_μ und Δ_Σ für den Kovarianztyp 3, 2 Kovariablen, 2 Zielgrößen, Identitätslink und Design 2

von 4 Kovariablen und 3 Zielgrößen ist in Abbildung 5.17 zu sehen, dass die Modelle M1, M2, M4 weiterhin mit wachsender Stichprobengröße besser werden. Bezüglich Δ_μ verhalten sich die Modelle M1 und M2 ab Stichprobengröße 60 ähnlich und bezüglich Δ_Σ liefert das Modell M4 etwas bessere Resultate als die Modelle M1 und M2.

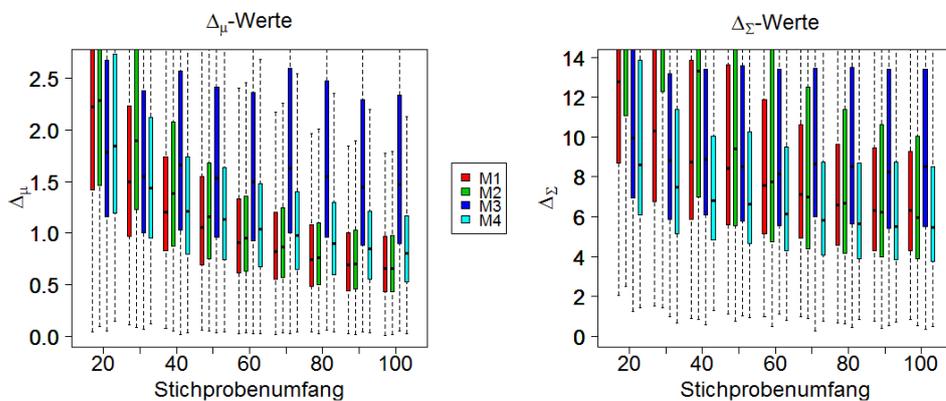


Abbildung 5.17: Boxplots der Distanzen Δ_μ und Δ_Σ für den Kovarianztyp 3, 4 Kovariablen, 3 Zielgrößen, Identitätslink und Design 2

Abbildung 5.18 zeigt die Ergebnisse bei Verwendung des Logarithmuslinks beispielhaft für 4 Kovariablen und 2 Zielgrößen. Die Schätzung basierend auf dem Modell M3 ver-

sagt auch hier, wie bereits bei den Kovarianztypen 1 und 2 beobachtet. Mit wachsendem Stichprobenumfang wird die Schätzung des Erwartungswertvektors für die Modelle M1, M2 und M4 besser, wobei das Modell M1 am besten abschneidet. Bezüglich der Kovarianzmatrix liegt ebenfalls das Modell M2 vorn.

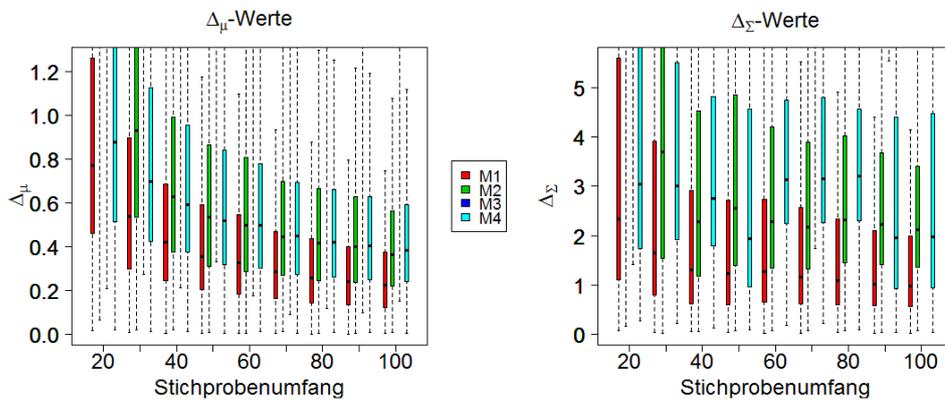


Abbildung 5.18: Boxplots der Distanzen Δ_μ und Δ_Σ für den Kovarianztyp 3, 4 Kovariablen, 2 Zielgrößen, Logarithmuslink und Design 2

Zusammenfassung

Insgesamt lassen die Simulationsergebnisse zusammenfassend folgende Schlüsse zu:

1. Mit wachsendem Stichprobenumfang oder Wiederholungen werden Schätzungen erwartungsgemäß besser.
2. Das einfache Vorgehen M5 wird immer von mindestens einem Modell übertroffen.
3. Bei Verwendung des Logarithmuslinks versagt das nichtparametrische Modell M3 in der zuvor beschriebenen Simulationsstudie. Das kann möglicherweise durch eine bessere Wahl der Bandbreite behoben werden.
4. Das Modell M4 liefert bei Verwendung von Design 1 recht stabile Resultate, die immer besser sind im Vergleich zum einfachen Vorgehen M5. Der Erwartungswert

wird durch die Modellannahme im heuristischen Modell in Kapitel 5.4 oft gut geschätzt und hat somit einen Vorteil gegenüber dem nichtparametrischen Modell M3.

5. Bei Versuchen ohne Wiederholung liefert das Modell M2 gerade bei geringem Datenumfang oft die schlechtesten Schätzungen. Dies kann in einigen Fällen möglicherweise auf Konvergenzprobleme zurückgeführt werden, die unter Umständen etwa durch multiple Starts verhindert werden können. Es kann jedoch sein, dass die Stichprobengröße zu klein ausfällt oder die Datenqualität unzureichend ist, wie etwa die Ergebnisse in Abbildung 5.3 demonstrieren.
6. Es gibt keinen klaren Favoriten, der in jedem Fall die beste Schätzung liefert. Allerdings bieten die vorgestellten Modelle eine geeignete Auswahl. Bei einer vorliegenden Datensituation kann ein geeignetes Modell aus der Menge der in Kapitel 5 vorgestellten Ansätze ausgewählt und das beste Modell beispielsweise anhand von geeigneten Kriterien (wie etwa dem AIC) bestimmt werden.

6. Eigenschaften und Implementierung der JOP-Methode

Basierend auf den in Kapitel 4 vorgestellten doppelt generalisierten linearen Modellen bei unkorrelierten Zielgrößen oder in Kapitel 5 präsentierten Modellen für den Erwartungswertvektor und die Kovarianzmatrix bei korrelierten Zielgrößen kann eine robuste Prozessauslegung erfolgen. Dies kann mithilfe der in Kapitel 3.3 erwähnten JOP-Methode erfolgen, bei der die geschätzte Risikofunktion 3.7

$$\hat{R}(x) = \text{Spur}(C\hat{\Sigma}(x)) + (\hat{\mu}(x) - \tau)^T C (\hat{\mu}(x) - \tau)$$

für eine ganze Folge von Kostenmatrizen C_1, \dots, C_m minimiert wird. In diesem Kapitel wird näher auf die JOP-Methode eingegangen. Zudem wird die Pareto-Optimalität nachgewiesen. Abschließend wird das R-Paket JOP (Kuhnt und Rudak, 2013b) anhand eines Beispiels vorgestellt. Teile des Kapitels wurden bereits in Erdbrügge et al. (2011) veröffentlicht.

In diesem Kapitel wird davon ausgegangen, dass die einzelnen Zielvariablen $Y_i|x_i$, $i = 1, \dots, p$, unkorreliert sind.

Wahl der Kostenmatrix

Eine sinnvolle Wahl der Kostenmatrix C wird nachfolgend gemessen an zwei Kriterien (siehe auch Kuhnt und Erdbrügge (2004)). Zum einen soll der Verlust immer positiv und zum anderen muss der Verlust skaleninvariant sein, das heißt, dass verschiedene Maßeinheiten wie cm oder mm zum gleichen Verlustwert führen müssen. Die erste Bedingung wird erfüllt durch die Wahl von positiv definiten Kostenmatrizen. Hier werden insbesondere diagonale positiv definite Kostenmatrizen betrachtet. Des Weiteren soll der Verlust invariant sein gegenüber linearen Transformationen, wie etwa $\tilde{Y} = V \cdot Y$ für eine beliebige Diagonalmatrix $V \in \mathbb{R}^{n \times n}$. Um diesen beiden Zielen gerecht zu werden, wird

die Kostenmatrix weiter spezifiziert durch eine Standardisierungsmatrix A_Y und eine Gewichtsmatrix W über

$$C_Y = A_Y^T W A_Y. \quad (6.1)$$

Die Gewichtsmatrix soll dabei dazu dienen, die Bedeutung der Erreichung des Zielwertes für die einzelnen Zielgrößen unterschiedlich zu gewichten. Hingegen soll die Standardisierungsmatrix die Invarianz gegenüber linearen Transformationen sicherstellen.

Wahl der Standardisierungsmatrix

Wird die lineare Transformation $\tilde{Y} = VY$ sowie der entsprechend transformierte Zielwertvektor $\tilde{\tau} = V\tau$ zunächst in die Verlustgleichung (3.6) eingesetzt, ergibt sich

$$\begin{aligned} \text{loss}(\tilde{Y}) &= (\tilde{Y} - \tilde{\tau})^T C_{\tilde{Y}} (\tilde{Y} - \tilde{\tau}) \\ &= (V \cdot Y - V \cdot \tau)^T C_{\tilde{Y}} (V \cdot Y - V \cdot \tau) \\ &= (Y - \tau)^T V^T C_{\tilde{Y}} V (Y - \tau). \end{aligned}$$

Damit der Verlust invariant ist gegenüber der linearen Transformation $\tilde{Y} = V \cdot Y$, muss folglich nach Einsetzen von (6.1) gelten

$$\begin{aligned} V^T C_{\tilde{Y}} V &= V^T A_{\tilde{Y}}^T W A_{\tilde{Y}} V \\ &\stackrel{!}{=} C_Y = A_Y^T W A_Y. \end{aligned}$$

Damit ergibt sich die folgende Bemerkung.

Bemerkung 6.1 Sei $Y \in \mathbb{R}^n$ ein Zielgrößenvektor, $V \in \mathbb{R}^{n \times n}$ eine Diagonalmatrix und $\tilde{Y} = VY$ eine lineare Transformation. Dann ist $\text{loss}(\tilde{Y}) = \text{loss}(Y)$ genau dann, wenn

$$V^T A_{\tilde{Y}}^T = A_Y^T \quad (6.2)$$

gilt.

Bezeichne $\hat{E}(Y_r|x)$ die geschätzte Funktion des Erwartungswerts $E(Y_r|x)$ und $\widehat{\text{Var}}(Y_r|x)$ die geschätzte Funktion der Varianz $\text{Var}(Y_r|x)$ der Zielgröße Y_r abhängig von x . Zwei mögliche datenbasierte Standardisierungsmatrizen sind

$$A_Y^1 = \text{diag} \left(\left[\frac{1}{m} \sum_{i=1}^m \widehat{\text{Var}}(Y_r|x^i) \right]_{r=1,\dots,p}^{-\frac{1}{2}} \right) \quad (6.3)$$

und

$$A_Y^2 = \text{diag} \left(\left[\frac{1}{m} \sum_{i=1}^m \hat{E}(Y_r|x^i) \right]_{r=1,\dots,p}^{-1} \right). \quad (6.4)$$

Hier ist m die Anzahl der Designpunkte $x^i = (x_1^i, \dots, x_k^i)^T$ ($i = 1, \dots, m$) im Versuchsplan. Der Punkt x^i wird dabei n_i -mal wiederholt. Damit ist $n = \sum_{i=1}^m n_i$ die Gesamtzahl der Beobachtungen. Beide Matrizen A_Y^1 und A_Y^2 erfüllen Bedingung 6.2. Die Ausdrücke $\widehat{\text{Var}}(Y_r|x^i)$ und $\hat{E}(Y_r|x^i)$ können etwa auch bei Wiederholungen des Versuchsplans durch die Stichprobenvarianz und das arithmetische Mittel ersetzt werden.

Die Standardisierungsmatrizen A_Y^1 und A_Y^2 gewährleisten die Invarianz der Verlustfunktion gegenüber linearen Transformationen. Die Matrix A_Y^1 erfüllt im Gegensatz zu A_Y^2 sogar die Invarianz gegenüber affin linearen Transformationen der Art $\tilde{Y} = VY + b$ mit $b \in \mathbb{R}^p$. Damit ist der Verlust mit der Wahl von A_Y^1 sowohl skalen- als auch translationsinvariant. Translationsinvarianz des Verlustes ist für die Matrix A_Y^2 nicht erfüllt. Ein Beispiel für Skaleninvarianz sind etwa unterschiedliche Einheiten wie cm oder mm . Hier sollte der Verlust immer gleich sein, egal welche Einheit man verwendet. Liegen die Zielgrößen und Zielwerte in einer standardisierten Form $\frac{Y_i - a}{b}$ und $\frac{\tau_i - a}{b}$ vor, so sichert die Matrix A_Y^1 denselben Verlust wie bei Verwendung von Y und τ . Daher ist die Verwendung der Standardisierungsmatrix A_Y^1 zu bevorzugen.

Wahl der Gewichtsmatrix

Eine konkrete Wahl einer Reihe von Gewichtsmatrizen $(W_t)_{t=1,\dots,T}$ muss sicherstellen, dass die Gewichtsmatrizen paarweise linear unabhängig sind. Andernfalls würde die Mi-

Minimierung der Risikofunktion für unterschiedliche Kostenmatrizen zu gleichen Lösungen führen. Dazu sei $d \in \mathbb{R}^p$ fest und $a = (a_1, \dots, a_T) \in \mathbb{R}^T$ bezeichne einen **Streckungsvektor**. Dann wird eine Folge von Gewichtsmatrizen wie folgt definiert (Erdbrügge und Kuhnt, 2002).

$$\log w_t = d \cdot a_t, t = 1, \dots, T \quad (6.5)$$

$$W_t = \text{diag}(w_t) \quad (6.6)$$

Damit ist jede Gewichtsmatrix festgelegt durch den fest gewählten Vektor d und einen Eintrag des Streckungsvektors a . Der log-Ausdruck garantiert hier zum einen, dass die Diagonaleinträge der Gewichtsmatrix positiv sind. Zum anderen sind so zwei unterschiedliche Gewichtsmatrizen paarweise linear unabhängig und es ergeben sich unterschiedliche Lösungen bei Minimierung der Risikofunktion.

6.1. Pareto-Optimalität der JOP-Methode

Im eindimensionalen Raum lässt sich für $x, y \in \mathbb{R}$ eindeutig entscheiden, ob $x \leq y$ oder $x \geq y$ gilt. Im n -dimensionalen Raum \mathbb{R}^n ist dies im Allgemeinen nicht mehr so möglich. Es gibt unterschiedliche Optimalitätskonzepte, wie etwa die Pareto Optimalität (Hillermeier, 2001). Hier wird von der folgenden Definition bezüglich der Minimierung ausgegangen. Analog kann ein Konzept der Pareto Optimalität bezüglich der Maximierung definiert werden.

Definition 6.2 (Pareto Optimalität) Sei \leq eine Ordnungsrelation im \mathbb{R}^p definiert durch

$$y \leq y^* \Leftrightarrow y^* - y \in \mathbb{R}_+^p \text{ für } y, y^* \in \mathbb{R}^p,$$

wobei \mathbb{R}_+^n den nicht-negativen Orthanten von \mathbb{R}^n bezeichnet. Gegeben sei eine vektorwertige Funktion $f : X \subset \mathbb{R}^k \rightarrow \mathbb{R}^p$. Ein Punkt $y^* \in f(X)$ wird effizient genannt bezüglich der Ordnungsrelation \leq genau dann, wenn kein $y \in f(X)$, $y \neq y^*$, existiert mit $y \leq y^*$.

Ein Punkt $x^* \in \mathbb{R}^k$ mit $y^* = f(x^*)$ heißt **Pareto optimal** genau dann, wenn y^* effizient ist.

Für zwei Punkte $y_1 = \begin{pmatrix} -1 \\ -2 \end{pmatrix}$ und $y_2 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ gilt zum Beispiel

$$y_2 - y_1 = \begin{pmatrix} 2 \\ 2 \end{pmatrix} \in \mathbb{R}_+^2$$

und nach Definition 6.2 somit $y_1 \leq y_2$. Dagegen gilt für $y_1 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ und $y_2 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ weder $y_1 \leq y_2$ noch $y_2 \leq y_1$.

Ein Punkt $x \in \mathbb{R}^k$ mit $y = f(x)$ ist nicht Pareto optimal, wenn ein Punkt x^* mit $y^* = f(x^*)$ existiert mit $y^* \leq y$ und mindestens einem Index $i_0 \in \{1, \dots, p\}$ mit $y_{i_0}^* < y_{i_0}$.

Die JOP-Methode liefert Pareto optimale Punkte für das folgende multikriterielle Optimierungsproblem

$$\min_{x \in X} R_{vec}(x), \quad (6.7)$$

wobei $R_{vec}(x) = (r_1(x), \dots, r_p(x))^T \in \mathbb{R}^p$ mit $r_i(x) = \sigma_i^2(x) + (\mu_i(x) - \tau_i)^2$ für $i = 1, \dots, p$ ist. Dieses nachfolgend bewiesene Ergebnis wurde bereits in Erdbrügge et al. (2011) publiziert.

Satz 6.3 Gegeben sei eine Risikofunktion $R(x)$ aus (3.7) mit $x \in X \subset \mathbb{R}^k$ und $C \in \mathbb{R}^{p \times p}$ eine positiv definite diagonale Kostenmatrix. Dann ist jede Lösung $x^* = \arg \min_{x \in X} R(x)$ Pareto optimal für $R_{vec}(x^*)$.

Beweis:

Im Falle einer diagonalen Kostenmatrix kann die Risikofunktion geschrieben werden als

$$\begin{aligned} R(x) &= \text{Spur}(C\Sigma(x)) + (\mu(x) - \tau)^T C (\mu(x) - \tau) \\ &= \sum_{i=1}^p c_i r_i(x) \end{aligned}$$

mit $r_i(x) = \sigma_i^2(x) + (\mu_i(x) - \tau_i)^2$ und $c_i = [C]_{ii}$. Der Beweis erfolgt durch Widerspruch. Angenommen, dass eine Lösung $x^* = \arg \min_{x \in X} R(x)$ mit $y^* := R_{vec}(x^*) \in \mathbb{R}^p$ nicht Pareto optimal ist. Dann existiert mindestens ein $x \in \mathbb{R}^k$ mit $y = R_{vec}(x) \in \mathbb{R}^p$, so dass gilt

$$y_i \leq y_i^* \quad \forall \quad i = 1, \dots, k$$

und für mindestens ein j gilt

$$0 \leq y_j < y_j^*.$$

Wegen $\sigma_i^2(x) > 0$, $(\mu_i(x) - \tau_i)^2 \geq 0$ und $c_i > 0$ (C positiv definite Diagonalmatrix) gilt $0 \leq y_i \quad \forall \quad i = 1, \dots, p$. Damit folgt

$$\begin{aligned} R(x^*) &= \sum_{i=1}^p c_i r_i(x^*) = \sum_{\substack{i=1 \\ i \neq j}}^p c_i r_i(x^*) + c_j r_j(x^*) \\ &= \sum_{\substack{i=1 \\ i \neq j}}^p c_i y_i^* + c_j y_j^* \geq \sum_{\substack{i=1 \\ i \neq j}}^p c_i y_i + c_j y_j^* \\ &> \sum_{\substack{i=1 \\ i \neq j}}^p c_i y_i + c_j y_j = \sum_{\substack{i=1 \\ i \neq j}}^p c_i r_i(x) + c_j r_j(x) \\ &= \sum_{i=1}^p c_i r_i(x) = R(x) \geq 0. \end{aligned}$$

Damit ergibt sich $R(x^*) > R(x)$, womit x^* die Risikofunktion nicht minimiert. Folglich ist x^* Pareto optimal. \square

Die JOP-Methode kann als Gewichtungsmethode für das multikriterielle Optimierungsproblem 6.7 angesehen werden mit

$$R(x) = \sum_{i=1}^p c_i \cdot (\sigma_i^2(x) + (\mu_i(x) - \tau_i)^2) \quad (6.8)$$

$$:= \sum_{i=1}^p c_i \cdot r_i(x), \quad (6.9)$$

wobei $r_i(x) = \sigma_i^2(x) + (\mu_i(x) - \tau_i)^2$ ist. Die Wahl der Kostenmatrix C hat folglich entscheidenden Einfluss auf die Lösung der Minimierung der Risikofunktion. Dies wird

besonders deutlich durch Betrachtung der Risikofunktion für den Fall einer diagonalen Kostenmatrix, wie in Satz 6.3. Das Optimierungsproblem

$$\min_{x \in X} \sum_{i=1}^p c_i \cdot r_i(x),$$

kann als ein lineares Optimierungsproblem

$$\min_{z \in B_R} c^T z,$$

im Bildraum $B_R = \{z \in \mathbb{R}^p \mid \exists x \in \mathbb{R}^k : z_i = r_i(x) \forall i \in 1, \dots, p\}$ der einzelnen Risiken r_i gesehen werden. Hier dient c als Normalenvektor der Hyperebene $c^T z = \alpha$. Die Minimierung wird in Abbildung 6.1 für den Fall von zwei Zielgrößen, d.h. $R_{vec}(x) = (r_1(x), r_2(x))^T$, deutlich. Die graue Menge stellt eine beispielhafte Bildmenge dar und das Bild der Pareto-Front ist hier grün (-) hervorgehoben. Die schwarzen parallelen Linien bezeichnen die Hyperebenen $c^T z = \alpha$ für verschiedene Wahlen von α . Schließlich folgt das Bild des Pareto optimalen Punktes (•) als Lösung für ein minimales α . Für jede Wahl

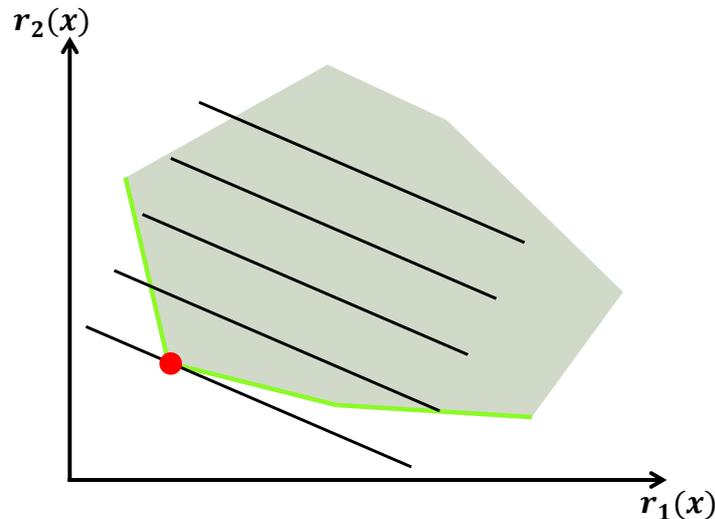


Abbildung 6.1: Grafische Interpretation zur Berechnung einer Pareto optimalen Lösung für eine fest gewählte Kostenmatrix C

der Kostenmatrix C wird ein anderer Pareto optimaler Punkt bestimmt, wie bereits in Satz 6.3 gezeigt wurde. Angemerkt werden muss dabei, dass bei Nichtkonvexität der

Bildmenge von R_{vec} nicht die ganze Pareto-Front ermittelt werden kann. Da die gesamte Pareto-Front im Allgemeinen überabzählbar ist und folglich aus verschiedenen solcher Kompromisse besteht, ist es von besonderem Interesse eine ganze Reihe von unterschiedlichen Pareto optimalen Punkten zu berechnen und anschließend anhand von geeigneten Kriterien oder prozessbasiertem Wissen einen geeigneten Kompromiss auszusuchen. Dies ist eine besondere Motivation für die Wahl einer ganzen Reihe von Kostenmatrizen.

6.2. Algorithmus und Implementierung im R-Paket JOP

Alle Schritte der JOP-Methode sind in folgendem Algorithmus zusammengefasst.

Algorithmus 6.4 (JOP für unkorrelierte Zielgrößen)

1. Festlegung von Zielwerten für jede einzelne Zielgröße.
2. Anpassung doppelt generalisierter linearer Modelle für jede einzelne Zielgröße.
3. Spezifizierung einer Folge von Kostenmatrizen nach (6.1) und Wahl der Gewichtsmatrizen nach (6.5).
4. Minimierung der Risikofunktion für jede Kostenmatrix C_t , $t = 1, \dots, T$.
5. Darstellung der Ergebnisse mithilfe des Joint Optimization Plots.
6. Wahl eines geeigneten Kompromisses auf Grundlage von Sachkenntnis oder mithilfe eines Kriteriums, z.B.

$$x^* = \operatorname{argmin} \sum_i^p ([Cov(Y|x)]_{ii} + (E(Y_i|x) - \tau_i)^2). \quad (6.10)$$

Dieser Algorithmus ist im R-Paket **JOP** implementiert und auf CRAN verfügbar (Kuhnt und Rudak, 2013a,b). Es ermöglicht die multikriterielle Optimierung multipler Zielgrößen auf Grundlage eines statistischen Versuchsplans.

Das Paket **JOP** besteht aus den folgenden Komponenten.

- `JOP`: Dies ist die Hauptfunktion des Pakets **JOP** und ermöglicht eine automatische Modellwahl doppelt generalisierter linearer Modelle (siehe Kapitel 4.2) für jede Zielgröße, die Berechnung der Kostenmatrizen sowie die Schätzung und Minimierung der Risikofunktion für jede Kostenmatrix. Alternativ können Funktionen für die Schätzung der Erwartungswerte und Varianzen der einzelnen Zielgrößen direkt übergeben werden.
- `plot.JOP`: Diese Funktion erzeugt einen Joint Optimization Plot (siehe Abbildung 6.2), der die Ergebnisse der Minimierung der Risikofunktion grafisch darstellt.
- `locate`: Mithilfe dieser Funktion kann der Anwender per Mausklick eine optimale Lösung auswählen und erhält die dazugehörigen Kovariablenwerte.

Nachfolgend wird die JOP-Methode anhand eines exemplarischen Datensatzes bestehend aus den im Paket enthaltenen Objekten `datax` und `datay` dargestellt. Dabei enthält `datax` einen Versuchsplan mit 2 Kovariablen X_1 und X_2 und `datay` die dazugehörigen beobachteten Werte zweier Zielvariablen Y_1 und Y_2 .

Zunächst werden doppelt generalisierte lineare Modelle für die beiden Zielgrößen geschätzt. Dazu wird die in JOP enthaltene automatische Modellselektion genutzt. Dabei wird zuerst für jede Zielgröße ein Erwartungswertmodell mit konstanter Dispersion durch eine Kombination aus Vorwärts- und Rückwärtsselektion bestimmt. Das maximale Modell enthält dabei Haupteffekte, Wechselwirkungen und quadratische Effekte. Anschließend werden doppelt generalisierte lineare Modelle für jede Zielgröße bestimmt. Dazu werden die bereits ermittelten Effekte für die Erwartungswertmodelle verwendet und das maximale Modell für die Dispersion enthält alle Haupteffekte, Wechselwirkungen und quadratischen Effekte. Daraufhin wird iterativ in jedem Schritt der Effekt im Dispersionsmodell mit dem kleinsten Wert $t_i = \frac{\hat{\beta}_i}{\sqrt{\text{Var}\hat{\beta}_i}}$ solange entfernt, bis für alle Werte $|t_i| > z_{1-\frac{\alpha}{2}}$ gilt, wobei $z_{1-\frac{\alpha}{2}}$ das $(1 - \frac{\alpha}{2})$ -Quantil der Standardnormalverteilung ist.

Hier wird $\alpha = 0.05$ gewählt. Die geschätzten Modelle lauten:

$$\begin{aligned}\widehat{E}(Y_1|x_1, x_2) &= 26.74 + 2.87x_1 - 11.49x_2 & (6.11) \\ \widehat{Var}(Y_1|x_1, x_2) &= \exp(3.85 + 0.48x_2) \\ \widehat{E}(Y_2|x_1, x_2) &= 0.06 + 0.01x_2 \\ \widehat{Var}(Y_2|x_1, x_2) &= \exp(-9.13).\end{aligned}$$

Die geschätzte Varianz für Y_1 hängt somit von der Kovariable x_2 ab.

Die zu erreichenden Zielwerte lauten bei dem obigen Beispiel 0 für Y_1 und 0.05 für Y_2 . Der Vektor der Zielwerte τ wird als Liste übergeben, wobei sich der i-te Eintrag auf die i-te Zielgröße bezieht. Ein numerischer Zielwert steht für einen bestimmten Wert, der erzielt werden soll. Wird der Zielwert "min"(oder "max") übergeben, so berechnet JOP zunächst für jede Zielgröße den kleinstmöglichen (oder größtmöglichen) Wert auf Basis der geschätzten oder übergebenen Modelle für jede Zielgröße und setzt diesen intern als Zielwert. Der Minimalaufruf

```
JOP(datax = datax, datay = datay, tau = list(0, 0.05))
```

berechnet 10 Kostenmatrizen und minimiert für jede Kostenmatrix die Risikofunktion. Das JOP-Objekt `out` enthält unter anderem die berechneten optimalen Zielgrößenwerte und zugehörigen Kovariablenwerte. Eine komplette Liste der zu übergebenden Argumente sowie der in dem JOP-Objekt enthaltenen Komponenten findet man im Anhang in Tabelle A.1 und A.2. Der Befehl `plot(out1)` liefert den zugehörigen Joint Optimization Plot in Abbildung 6.2. Der Anwender kann grafisch einen geeigneten Kompromiss mithilfe eines Mausklicks auswählen. In Abbildung 6.2 wird deutlich, dass die Zielgrößen ihre Zielwerte nicht gleichzeitig erreichen können. Während die Werte für Y_1 auf der rechten Seite der rechten Grafik sich dem Zielwert nähern, entfernen sich die Werte für Y_2 von ihrem Zielwert. Analoges gilt für die linke Seite der Grafik. Daher muss ein geeigneter Kompromiss gefunden werden. Einen möglichen Kompromiss kann man beispielsweise in der Mitte der rechten Grafik finden, wo beide Zielwerte simultan möglichst

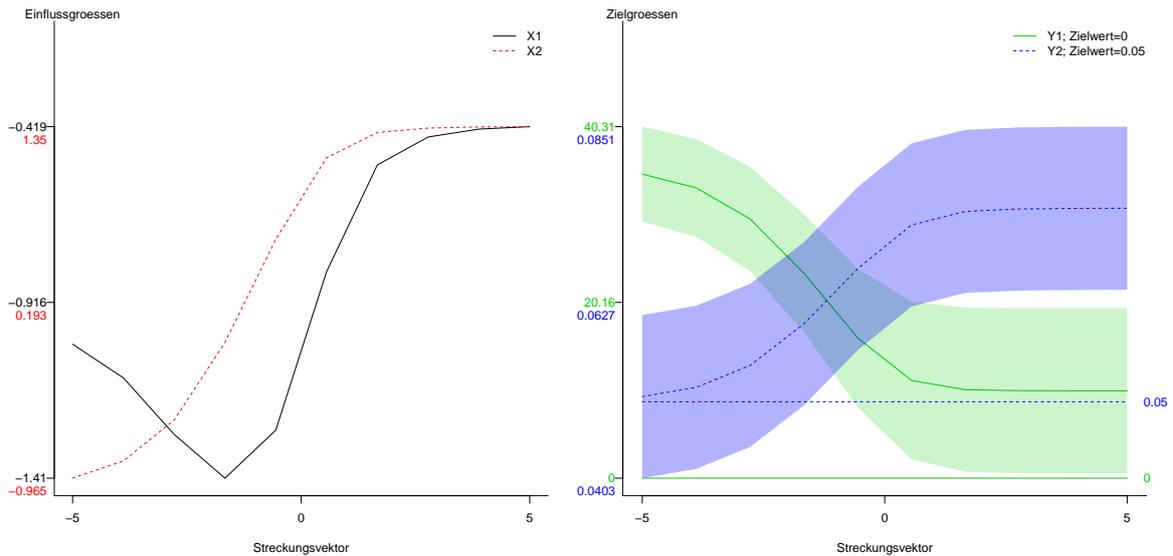


Abbildung 6.2: Beispielhafter Joint Optimization Plot

nahe am Zielwert liegen. Diesen Kompromiss erhält man, indem man mit der Maus nach Ausführung des Befehls `locate(out1)` auf die bevorzugte Stelle auf der rechten Grafik klickt. Ein möglicher Kompromiss ist beispielhaft in Abbildung 6.3 dargestellt. Zudem können die üblichen grafischen Parameter an die `plot`- und `locate`-Funktion übergeben werden.

Der Anwender kann entweder die in JOP enthaltene automatische Modellselektion nutzen oder alternativ eigene Modelle auf unterschiedliche Art übergeben, falls etwa schon Informationen über den Einfluss der Kovariablen auf die Zielgrößen vorliegen. So kann der Anwender eine Liste von `formula`-Objekten für den Erwartungswert und die Dispersion jeder Zielgröße spezifizieren mit

```
form.mean <- list(as.formula(Y1 ~ X1 + X2),
+ as.formula(Y2 ~ X2))
form.disp <- list(as.formula(d ~ X_2), as.formula(d ~ 1))
```

und anschließend über den Befehl

```
out2 <- JOP(datax = datax, datay = datay, tau = list(0, 0.05),
+ form.mean = form.mean, form.disp = form.disp, joplot=TRUE).
```

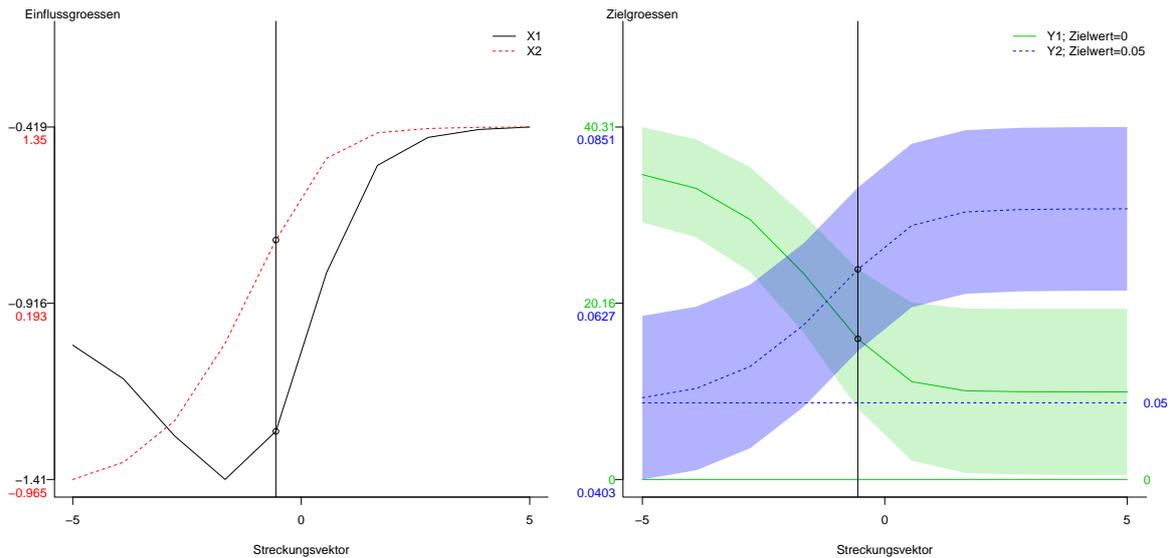


Abbildung 6.3: Beispielhafte Kompromisswahl im Joint Optimization Plot für out1

an JOP übergeben. Auf diese Weise erhält man auch den Joint Optimization Plot in Abbildung 6.2. Des Weiteren können auch konkrete Modelle direkt an JOP in Form von R-Funktionen über Listen `mean.model=list(mean1fun,mean2fun)` und `var.model=list(var1fun,var2fun)` übergeben werden. Die R-Funktion `meanfun1` für $E(Y_1|x_1, x_2)$ in Gleichung (6.11) sieht beispielsweise wie folgt aus:

```
meanfun1 <- function(x)
return(26.74+2.87x[1]-11.49x[2])
```

Mithilfe des Befehls

```
out3 <- JOP(datax = datax, datay = datay, tau = list(0, 0.05),
+ mean.model = mean.model, var.model=var.model,joplot=TRUE).
```

erhält man wieder den Joint Optimization Plot in Abbildung 6.2.

Für weitere Details zum R-Paket JOP wird auf Kuhnt und Rudak (2013a) und Kuhnt und Rudak (2013b) verwiesen. Eine mögliche Anwendung der JOP-Methode ist zu finden in Rehage et al. (2012a).

7. Erweiterung der JOP-Methode auf multiple korrelierte Zielgrößen

Um mögliche Korrelationen zwischen den Zielgrößen einzubeziehen, wird in diesem Kapitel die JOP-Methode zum einen durch die Verwendung multivariater Regressionsmodelle (siehe Kapitel 5) erweitert, die sowohl den Erwartungswertvektor als auch die Kovarianzmatrix mithilfe der Designvariablen modellieren. Zum anderen erfolgt die Erweiterung durch den Einsatz nichtdiagonaler Kostenmatrizen. Teile dieses Kapitels wurden bereits in Rudak et al. (2015a) publiziert.

Die Konstruktion nichtdiagonaler Kostenmatrizen erfolgt als natürliche Erweiterung der Wahl in Kapitel 6. Für $t \in \{1, \dots, T\}$ wird die Kostenmatrix C_t erneut aufgeteilt in

$$C_t = A_Y^T W_t A_Y, \quad (7.1)$$

wobei A_Y eine Standardisierungsmatrix der Form (6.3) oder (6.4) darstellt. Des Weiteren bezeichne $W_0 \in \mathbb{R}^{p \times p}$ eine symmetrische reelle Matrix und $(a_t)_{t \in \{1, \dots, T\}} \subset \mathbb{R}$ sei eine Folge von reellen Werten mit $a_i \neq a_j$ für alle $i, j = \{1, \dots, m\}$. Als nächstes wird eine nichtdiagonale Gewichtsmatrix W_t definiert durch

$$W_t = \exp(a_t \cdot W_0), \quad (7.2)$$

wobei $\exp(\cdot)$ die exponentielle Matrixtransformation (Bellman, 1970)

$$\exp(A) = \sum_{i=0}^{\infty} \frac{A^i}{i!} \quad (7.3)$$

meint. Für eine symmetrische Matrix $A \in \mathbb{R}^{p \times p}$ ist $\exp(A)$ positiv definit und symmetrisch wegen:

1. Sind $\lambda_i > 0$ für $i = 1, \dots, p$ positive Eigenwerte von A , so sind $\exp(\lambda_i) > 0$ positive Eigenwerte von $\exp(A)$.
2. Sei $A = PDP^T$ die Spektralzerlegung von A mit einer orthonormalen Matrix P und einer Diagonalmatrix D , so gilt $\exp(A) = P \exp(D) P^T = (P \exp(D) P^T)^T = (\exp(A))^T$.

Eine geeignete Wahl der Kostenmatrizen sollte garantieren, dass für $i, j \in \{1, \dots, m\}$ mit $i \neq j$ immer

$$C_i \neq a \cdot C_j \quad \forall a \in \mathbb{R} \quad (7.4)$$

gilt. Falls die Bedingung (7.4) für zwei Matrizen C_i und C_j nicht gilt, so unterscheiden sich die zugehörigen Risikofunktionen lediglich durch Multiplikation mit einer Konstante und damit erhält man durch Minimierung dieselben optimalen Lösungen. Der folgende Satz zeigt, dass durch die Wahl der Gewichtsmatrizen der Form (7.2) die Bedingung (7.4) erfüllt wird.

Satz 7.1 *Sei $W_0 \in \mathbb{R}^{p \times p}$ eine symmetrische Matrix mit mindestens zwei unterschiedlichen Diagonaleinträgen und sei $A_Y \in \mathbb{R}^{p \times p}$ eine beliebige reguläre Standardisierungsmatrix. Des Weiteren sei $(a_t)_{t \in \{1, \dots, T\}} \subset \mathbb{R}$ eine Folge von reellen Werten mit $a_i \neq a_j \forall i, j \in \{1, \dots, T\}$. Dann ist für die Folge von Kostenmatrizen W_1, \dots, W_T entsprechend der Gleichung (7.2) die Bedingung (7.4) erfüllt.*

Beweis:

Die Kostenmatrizen C_t werden nach Konstruktion aufgeteilt in $C_t = A_Y^T W_t A_Y$ (siehe (7.1)) mit einer konstanten von t unabhängigen Standardisierungsmatrix A_Y und Gewichtsmatrizen W_t entsprechend der Gleichung (7.2) für $t = 1, \dots, m$.

Der Beweis erfolgt durch Widerspruch. Angenommen, es gilt $\exp(a_i \cdot W_0) = b \cdot \exp(a_j \cdot W_0)$ für bestimmte Indizes $i \neq j \in \{1, \dots, T\}$ und eine Konstante $b \neq 1 \in \mathbb{R}$. Wegen der Symmetrie der reellen Matrix W_0 existiert die Spektralzerlegung $W_0 = T D T^T$ mit einer orthonormalen Matrix $T \in \mathbb{R}^{p \times p}$ und einer diagonalen Matrix $D \in \mathbb{R}^{p \times p}$ mit Diagonalelementen $\lambda_1, \dots, \lambda_p \in \mathbb{R}$. Mit dieser Spektralzerlegung folgt

$$\begin{aligned} \exp(a_i \cdot W_0) &= \sum_{j=0}^{\infty} \frac{(a_i W_0^j)}{j!} \\ &= T \left(\sum_{j=0}^{\infty} \frac{(a_i D)^j}{j!} \right) T^T \\ &= T \tilde{D}_i T^T, \end{aligned}$$

wobei $\tilde{D}_i \in \mathbb{R}^{p \times p}$ eine Diagonalmatrix mit Diagonaleinträgen $e^{a_i \lambda_r}$ für $r = 1, \dots, p$ ist. Analog ergibt sich $\exp(a_j \cdot W_0) = T \tilde{D}_j T^T$, wobei $\tilde{D}_j \in \mathbb{R}^{p \times p}$ wieder eine Diagonalmatrix ist mit den Diagonaleinträgen $e^{a_j \lambda_r}$ für $r = 1, \dots, p$. Aufgrund der Annahme $\exp(a_i \cdot W_0) = b \cdot \exp(a_j \cdot W_0)$ folgt $\frac{e^{a_i \lambda_r}}{e^{a_j \lambda_r}} = \text{const} = b$ für $r = \{1, \dots, p\}$. Damit ergibt sich durch Anwendung des Logarithmus

$$\lambda_r = \frac{\log(b)}{a_i - a_j} \quad \forall r \in \{1, \dots, p\}$$

und es folgt $\lambda := \lambda_1 = \dots = \lambda_p$. Deswegen folgt $W_0 = T D T^T = \lambda T T^T = \lambda I_{p \times p}$ mit der Identitätsmatrix $I_{p \times p}$, was einen Widerspruch zu der Annahme darstellt, dass W_0 eine Matrix ist, die mindestens zwei unterschiedlichen Werten auf der Diagonalen enthält. \square

Der Satz 7.1 schließt auch die Wahl der diagonalen Kostenmatrizen in Kapitel 6 ein. Damit zeigt der Satz 7.1, dass die Wahl der Folge von Kostenmatrizen sinnvoll ist im Sinne von (7.4). Aufgrund der erfüllten Bedingung (7.4) erhält man im Allgemeinen unterschiedliche Lösungen für alle Kostenmatrizen und kann daraus einen geeigneten Kompromiss auswählen. Aufgrund der Tatsache, dass die Minimierung der Risikofunktion für Kostenmatrizen, die sich durch Multiplikation mit einer Konstanten unterscheiden, zu identischen Ergebnissen führen, kann es sinnvoll sein die Kostenmatrizen zu standardisieren. Damit werden auch die Werte der Risikofunktion für unterschiedliche Kostenmatrizen aufgrund einer gemeinsamen oberen Schranke vergleichbar. Diese obere Schranke wird in der folgenden Behauptung gezeigt.

Satz 7.2 Sei $C = A_Y^T W A_Y \in \mathbb{R}^{p \times p}$ eine Kostenmatrix mit einer positiv definiten Matrix $W \in \mathbb{R}^{p \times p}$ mit Eigenwerten $0 \leq \lambda_1, \dots, \lambda_p \leq 1$ und einer invertierbaren Diagonalmatrix $A_Y \in \mathbb{R}^{p \times p}$. Seien weiterhin Y und X Zufallsvektoren und $\tau \in \mathbb{R}^p$. Die Risikofunktion

$$R(x) = \text{Spur}(C \Sigma(x)) + (\mu(x) - \tau)^T C (\mu(x) - \tau) \quad (7.5)$$

mit $\mu(x) = E(Y|X = x)$ und $\Sigma(x) = \text{Cov}(Y|X = x)$ ist beschränkt durch

$$0 \leq R(x) \leq \sum_{i=1}^p \left[\tilde{\Sigma}(x) \right]_{ii} + [\tilde{\mu}(x)]_i^2 \quad (7.6)$$

mit $\tilde{\Sigma}(x) = A_Y \Sigma(x) A_Y^T$ und $\tilde{\mu}(x) = A_Y \cdot (\mu(x) - \tau)$.

Beweis:

Unter Verwendung der Spektralzerlegung der Matrix $W = TDT^T$, wobei D eine Diagonalmatrix mit positiven Eigenwerten $0 \leq \lambda_1, \dots, \lambda_p \leq 1$ auf der Diagonalen ist, ergibt sich für die Risikofunktion

$$\begin{aligned} R(x) &= \text{Spur}(C\Sigma(x)) + (\mu(x) - \tau)^T C (\mu(x) - \tau) \\ &= \text{Spur}(A_Y^T T D T^T A_Y \Sigma(x)) + (\mu(x) - \tau)^T A_Y^T T D T^T A_Y (\mu(x) - \tau) \\ &= \text{Spur}(D(A_Y^T T)^T \Sigma(x) A_Y^T T) + (T^T A_Y (\mu(x) - \tau))^T D T^T A_Y (\mu(x) - \tau). \end{aligned}$$

Setze nun $\tilde{\Sigma}(x) = (A_Y^T T)^T \Sigma(x) A_Y^T T$ und $\tilde{\mu}(x) = T^T A_Y (\mu(x) - \tau)$. Dann folgt

$$\begin{aligned} R(x) &= \text{Spur}(D\tilde{\Sigma}(x)) + (\tilde{\mu}(x))^T D \tilde{\mu}(x) \\ &= \sum_{i=1}^p \lambda_i \left([\tilde{\Sigma}(x)]_{ii} + [\tilde{\mu}(x)]_i^2 \right). \end{aligned}$$

Bezeichne $e_{(i)} \in \mathbb{R}^p$ den Einheitsvektor mit Eins an der i -ten Stelle und Null sonst und sei $\tilde{t}_{(i)} = A_Y^T T e_{(i)}$. Damit gilt wegen der positiven Definitheit der Kovarianzmatrix $\Sigma(x)$, dass der Diagonaleintrag

$$[\tilde{\Sigma}(x)]_{ii} = e_{(i)}^T \tilde{\Sigma}(x) e_{(i)} = (A_Y^T T e_{(i)})^T \Sigma(x) A_Y^T T e_{(i)} = (\tilde{t}_{(i)})^T \Sigma(x) \tilde{t}_{(i)} > 0$$

positiv ist. Der verbleibende Ausdruck $[\tilde{\mu}(x)]_i^2$ ist aufgrund der Quadrierung positiv. Wegen $[\tilde{\mu}(x)]_i^2 > 0$ und $[\tilde{\Sigma}(x)]_{ii} > 0$ für $i = 1, \dots, p$ sowie $0 < \max_{i=1, \dots, p} \lambda_i < 1$ folgt

$$\begin{aligned} 0 &\leq R(x) \leq \sum_{i=1}^p [\tilde{\Sigma}(x)]_{ii} + [\tilde{\mu}(x)]_i^2 \\ &= \text{Spur}(\tilde{\Sigma}(x)) + (\mu(x) - \tau)^T A_Y^T T T^T A_Y (\mu(x) - \tau) \\ &= \text{Spur}(T^T A_Y \Sigma(x) A_Y^T T) + (\mu(x) - \tau)^T A_Y^T A_Y (\mu(x) - \tau) \\ &= \text{Spur}(T T^T A_Y \Sigma(x) A_Y^T) + (A_Y (\mu(x) - \tau))^T (A_Y (\mu(x) - \tau)) \\ &= \text{Spur}(A_Y \Sigma(x) A_Y^T) + (A_Y (\mu(x) - \tau))^T (A_Y (\mu(x) - \tau)). \end{aligned}$$

Mit $\tilde{\Sigma}(x) = A_Y \Sigma(x) A_Y^T$ und $\tilde{\mu}(x) = A_Y \cdot (\mu(x) - \tau)$ ergibt sich die Behauptung. □

Die obere Schranke in Satz 7.2 hängt somit nur noch von der Standardisierungsmatrix A_Y ab und ist unabhängig von den gewählten Gewichtsmatrizen. Es ergibt sich folgendes Korollar für Kostenmatrizen C , deren Eigenwerte auch größer als 1 sein können.

Korollar 7.3 *Sei $C = A_Y^T W A_Y \in \mathbb{R}^{p \times p}$ eine Kostenmatrix mit einer positiv definiten Matrix $W \in \mathbb{R}^{p \times p}$ mit Eigenwerten $\lambda_1, \dots, \lambda_p \geq 0$ und einer invertierbaren Diagonalmatrix $A_Y \in \mathbb{R}^{p \times p}$. Die Risikofunktion (7.5) ist beschränkt durch*

$$0 \leq \frac{R(x)}{U_{\max}} \leq \sum_{i=1}^p [\tilde{\Sigma}(x)]_{ii} + [\tilde{\mu}(x)]_i^2 \quad (7.7)$$

mit $U_{\max} = \max_{i=1, \dots, p} \lambda_i$, $\tilde{\Sigma}(x) = A_Y^T \Sigma(x) A_Y$ und $\tilde{\mu}(x) = A_Y \cdot (\mu(x) - \tau)$.

Beweis:

Die Matrix $\frac{C}{U_{\max}}$ mit $U_{\max} = \max_{i=1, \dots, p} \lambda_i$ besitzt positive Eigenwerte zwischen 0 und 1. Damit ergibt sich die Behauptung unmittelbar mit Satz 7.2. \square

Falls die Gewichtsmatrix W diagonal ist, so sind die Einträge auf der Diagonalen von W Eigenwerte und die Standardisierung mithilfe der Konstanten U_{\max} bedeutet nach Korollar 7.3 die Division durch den größten Eintrag auf der Diagonalen der Gewichtsmatrix W .

Für verschiedene Lösungen x_1, \dots, x_m , die die Risikofunktion für Kostenmatrizen der Gestalt wie in Gleichung (7.2) mit einer fest gewählten Matrix W_0 minimieren, kann die Wahl einer geeigneten Lösung auf Basis des Joint Optimization Plots erfolgen. Dies kann gemeinsam mit dem Anwender auf Grundlage von prozessbasiertem Wissen erfolgen. Des Weiteren kann die Kompromisslösung auch algorithmisch über

$$x^* = \operatorname{argmin}_{x \in \{x_1, \dots, x_m\}} \sum_{j=1}^p [\tilde{\Sigma}(x)]_{jj} + [\tilde{\mu}(x)]_j^2 \quad (7.8)$$

ausgewählt werden. Wird die JOP-Methode für N verschiedene Folgen von Kostenmatrizen $C_j^{(i)} = A_Y^T \exp(a_j W_0^{(i)}) A_Y$ für $j = 1, \dots, m$ mit fest gewählten Matrizen $W_0^{(i)}$ für $i = 1, \dots, N$ durchgeführt, so kann die Auswahl eines Kompromisses ebenfalls nach Kriterium (7.8) erfolgen, da die obere Schranke in Satz 7.2 und Korollar 7.3 unabhängig von

der Wahl der Gewichtsmatrizen ist. Nachfolgend ist das Vorgehen in einem Algorithmus zusammengefasst.

Algorithmus 7.4 (Wahl eines geeigneten Kompromisses)

1. Seien $x^{i,j}$ Lösungen der Minimierung der Risikofunktion für Kostenmatrizen $C_j^{(i)}$ für $j = 1, \dots, m$ und $i = 1, \dots, N$ mit einer Standardisierungsmatrix A_Y .
2. Bestimme einen geeigneten Kompromiss nach Kriterium (7.8) über

$$x^* = \underset{x \in \{x^{i,j} | i=1, \dots, N; j=1, \dots, m\}}{\operatorname{argmin}} \operatorname{Spur}(A_Y \Sigma(x) A_Y^T) + (A_Y^T (\mu(x) - \tau))^T (A_Y^T (\mu(x) - \tau)).$$

Als Nächstes zeigt der folgende Satz einen Bezug zur Pareto-Optimalität für die erweiterte JOP-Methode für korrelierte Zielgrößen auf.

Satz 7.5 Sei $x^* = \underset{x \in X}{\operatorname{argmin}} R(x)$ Lösung der Minimierung der Risikofunktion

$$R(x) = \operatorname{Spur}(C \Sigma(x)) + (\mu(x) - \tau)^T C (\mu(x) - \tau) \quad (7.9)$$

mit einer positiv definiten Kostenmatrix $C = A_Y^T W A_Y \in \mathbb{R}^{p \times p}$ der Gestalt (7.1), die nicht notwendigerweise diagonal ist, einer positiv definiten Gewichtsmatrix $W \in \mathbb{R}^{p \times p}$ und einer invertierbaren diagonalen Standardisierungsmatrix $A_Y \in \mathbb{R}^{p \times p}$. Seien ferner $T \in \mathbb{R}^{p \times p}$ eine orthonormale Matrix und $D \in \mathbb{R}^{p \times p}$ eine Diagonalmatrix mit $W = T D T^T$. Dann ist x^* mit $R_{\operatorname{vec}}(x^*)$ Pareto optimal für

$$\min_{x \in X} \tilde{R}_{\operatorname{vec}}(x), \quad (7.10)$$

wobei $\tilde{R}_{\operatorname{vec}}(x) = (\tilde{r}_1(x), \dots, \tilde{r}_p(x))^T \in \mathbb{R}^p$ mit $\tilde{r}_i(x) = \tilde{\sigma}(x)_i^2(x) + (\tilde{\mu}_i(x) - \tilde{\tau}_i)^2$, $\tilde{\sigma}_i(x) = [(A_Y^T T)^T \Sigma(x) A_Y^T T]_{ii}$, $\tilde{\mu}(x) = T^T A_Y \mu(x)$ und $\tilde{\tau} = T^T A_Y \tau$ für $i = 1, \dots, p$.

Beweis:

Wegen der positiven Definitheit der Matrix W existiert eine Matrix $T \in \mathbb{R}^{p \times p}$ und eine Diagonalmatrix $D \in \mathbb{R}^{p \times p}$ mit Einträgen $\lambda_i := [D]_{ii}$, so dass $W = T D T^T$ gilt. Nach Einsetzen von $C = A_Y^T T D T^T A_Y$ in die Risikofunktion (7.9) und mit den Bezeichnungen

$\tilde{\sigma}_i = [(A_Y^T T)^T \Sigma(x) A_Y^T T]_{ii}$, $\tilde{\mu}(x) = T^T A_Y \mu(x)$ und $\tilde{\tau} = T^T A_Y \tau$ ergibt sich ähnlich wie bereits im Beweis von Satz 7.2

$$\begin{aligned}
R(x) &= \text{Spur}(C\Sigma(x)) + (\mu(x) - \tau)^T C (\mu(x) - \tau) \\
&= \text{Spur}(A_Y^T T D T^T A_Y \Sigma(x)) + (\mu(x) - \tau)^T A_Y^T T D T^T A_Y (\mu(x) - \tau) \\
&= \text{Spur}(D(A_Y^T T)^T \Sigma(x) A_Y^T T) + (T^T A_Y \mu(x) - T^T A_Y \tau)^T D (T^T A_Y \mu(x) - T^T A_Y \tau) \\
&= \text{Spur}(D\tilde{\Sigma}(x)) + (\tilde{\mu}(x) - \tilde{\tau})^T D (\tilde{\mu}(x) - \tilde{\tau}) \\
&= \sum_{i=1}^p \lambda_i \left([\tilde{\Sigma}(x)]_{ii} + [\tilde{\mu}(x) - \tilde{\tau}]_i^2 \right). \\
&:= \sum_{i=1}^p \lambda_i \tilde{r}_i(x).
\end{aligned}$$

Wegen der positiven Definitheit sind die Eigenwerte $\lambda_i > 0$ positiv. Wie im Beweis von Satz 7.2 ergibt sich $[\tilde{\Sigma}(x)]_{ii} > 0$ und $[\tilde{\mu}(x) - \tilde{\tau}]_i^2 > 0$. Insgesamt folgt $\lambda_i > 0$ und $\tilde{r}_i(x) > 0$ für alle $i = 1 \dots, p$. Der Rest des Beweises folgt nun analog zum Beweis von 6.3 über Widerspruch. \square

Ist die Kostenmatrix $W \in \mathbb{R}^{p \times p}$ und damit auch $C \in \mathbb{R}^{p \times p}$ diagonal, so sind die Werte auf der Diagonalen von W gleichzeitig die Eigenwerte und die zugehörige Transformationsmatrix T ist gerade die Einheitsmatrix. Damit folgt

$$\text{Spur}(D\tilde{\Sigma}(x)) = \text{Spur}(A_Y^T D A_Y \Sigma(x)) = \text{Spur}(C\Sigma(x))$$

und entsprechend

$$(\tilde{\mu}(x) - \tilde{\tau})^T D (\tilde{\mu}(x) - \tilde{\tau}) = (\mu(x) - \tau)^T C (\mu(x) - \tau).$$

Folglich ist der Satz 7.5 eine Verallgemeinerung des Satzes 6.3.

Bemerkung 7.6 (Beziehung zu Wünschbarkeiten) *Mit Blick auf die Darstellung $\tilde{R}(x) = \sum_{i=1}^p \lambda_i \tilde{r}_i(x)$ können Wünschbarkeiten $d_i = \exp(-\tilde{r}_i(x))$ definiert werden, ähnlich wie in (3.1). Die Maximierung*

$$\operatorname{argmax}_{x \in \mathcal{X}} D(x) = \operatorname{argmax}_{x \in \mathcal{X}} \prod_{i=1}^p d_i(x)$$

führt wegen $\tilde{R}(x) = -\log(D(x))$ zur derselben Lösung wie bei der Minimierung der Risikofunktion $R(x)$.

Um auch Korrelationen auf der Modellierungsebene einzubeziehen, können die in Kapitel 5 zu diesem Zweck weiterentwickelten Modelle verwendet werden. Das Kapitel 5 bietet eine Auswahl an unterschiedlichen Modellen, bei denen sowohl der Erwartungswertvektor als auch die Kovarianzmatrix von Kovariablen abhängen darf. Der nichtparametrische multivariate Nadaraya-Watson-Kernschätzer setzt keine Kenntnisse über die Struktur des Erwartungswertvektors und der Kovarianzmatrix voraus, ist jedoch schwierig zu interpretieren. Das Matrix-Logarithmus-Modell und das Kovarianzregressionsmodell nehmen dabei eine bestimmte Struktur der Kovarianzmatrix an, wobei die exponentielle Transformation beim Matrix-Logarithmus-Modell hochgradig nichtlinear ist, während das Kovarianzregressionsmodell ein einfacheres quadratisches Modell für die Kovarianzmatrix ansetzt und eine etwas leichtere Interpretierbarkeit der Parameter zulässt. Bei Verwendung dieser Modelle und einer nichtdiagonalen Kostenmatrix zur Minimierung der Risikofunktion werden somit Korrelationen berücksichtigt.

Die JOP-Methode für korrelierte Zielgrößen wird wie folgt zusammengefasst. Dabei bleiben Kernelemente wie die Minimierung der Risikofunktion und die Darstellung mithilfe des Joint Optimization Plots so wie in Kapitel 6 unverändert.

Algorithmus 7.7 (JOP-Methode für multiple korrelierte Zielgrößen)

1. Schätzung des Erwartungswertvektors $E(Y|x)$ und der Kovarianzmatrix $Cov(Y|x)$ durch die Verwendung multivariater Regressionsmodelle der Art wie in Kapitel 5.
2. Festlegung von Werten $a_1, \dots, a_T \in \mathbb{R}$ für ein $T \in \mathbb{N}$, einer Matrix $W_0 \in \mathbb{R}^{p \times p}$ und einer Standardisierungsmatrix $A_Y \in \mathbb{R}^{p \times p}$. Berechnung der Kostenmatrizen C_1, \dots, C_m entsprechend (7.2).
3. Schätzung der Risikofunktion

$$\tilde{R}(x) = \text{trace}(C Cov(Y|x)) + (E(Y|x) - \tau)^T C (E(Y|x) - \tau).$$

4. *Minimierung der geschätzten Risikofunktion $\hat{R}(x)$ für alle Kostenmatrizen.*
5. *Erzeugung des Joint Optimization Plot.*
6. *Wahl eines geeigneten Kompromisses auf Grundlage der Erfahrung oder mithilfe eines Kriteriums, z.B. mithilfe des Algorithmus 7.4.*

8. Anwendungsbeispiele

In diesem Kapitel werden zwei Anwendungsbeispiele vorgestellt. Dabei wird beispielhaft das Kovarianzregressionsmodell benutzt. Die erste Anwendung zielt darauf ab Spritzgussparameter zu finden, die zu einer Verbesserung der Ausgewogenheit von Plastikradabdeckungen bei Kraftfahrzeugen führen. Der entsprechende Datensatz wurde bereits in Hamada und Balakrishnan (1998) analysiert. Als weiteres Anwendungsbeispiel wird ein Datensatz resultierend aus einem HVOF-Spritzprozess (siehe Kapitel 2) vorgestellt, bei dem es darum geht Maschinenparameter zu finden, die zu einer gewünschten Oberflächenbeschichtung führen.

8.1. Wheel Cover Experiment

Bei diesem Experiment wird die Herstellung einer möglichst leichten und ausbalancierten Radabdeckung bei Kraftfahrzeugen betrachtet. Es werden Einstellungen von Spritzgussparametern gesucht, die zu einem möglichst geringen Gewicht (Y_1) und einem ruhigen Laufverhalten (Y_2) führen. Das Gewicht Y_1 soll dabei im Bereich (710,715) und die Balance Y_2 im Bereich (0.3,0.4) liegen. Die Einflussgrößen sind

- x_1 : mold temperature (Werkzeugtemperatur),
- x_2 : close time (Endzeit),
- x_3 : booster time (Dauer der Druckerhöhung),
- x_4 : plunger time (Druckzeit des Kolbens),
- x_5 : pack pressure (Pressdruck),
- x_6 : hold pressure (Nachdruck),
- x_7 : barrel (Maßeinheit Barrel).

Es wurden Versuche basierend auf einem teilfaktoriellen Versuchsplan (2^{7-4}) durchgeführt und jede Einstellung wurde fünf mal wiederholt. Der Datensatz ist zu finden in Tabelle A.3 im Anhang. Aufgrund vorliegender Wiederholungen können einzelne Re-

gressionsmodelle basierend auf empirischen Mittelwerten, Varianzen und Korrelationen bestimmt werden. In Chiao und Hamada (2001) werden lineare Modelle basierend auf den empirischen Mittelwerten, empirischen Varianzen und empirischen Kovarianzen erstellt. Dabei wurden lediglich die Variablen F_1, F_2, F_4, F_5 und F_7 einbezogen. Es wurde eine Lösung gesucht mit $F_i \in \{-1, +1\}$ für $i = 1, \dots, 5$. Kuhnt und Erdbrügge (2004) verwenden dieselben Modelle wie in Chiao und Hamada (2001), um Lösungen innerhalb des experimentellen Bereiches zu finden und die in Chiao und Hamada (2001) gefundene Lösung zu verbessern. Hier wurden Parametereinstellungen zugelassen, die auch außerhalb des experimentellen Bereiches liegen. Somit können die Kovariablen auch Werte außerhalb des Intervalls $[-1, 1]$ annehmen, was einen Vergleich mit den hier erhaltenen Lösungen daher schwer möglich macht. In Salmasnia und Bashiri (2015) wurden auf ähnliche Weise wie in Chiao und Hamada (2001) Modelle basierend auf den empirischen Mittelwerten, den empirischen Varianzen und den empirischen Kovarianzen erstellt. Diese Modelle enthalten alle Kovariablen F_1, \dots, F_7 . Anschließend wurde eine Lösung basierend auf einer wünschbarkeitsbasierten Methode berechnet. Die erhaltenen optimalen Kovariablen werden im Folgenden als Vergleich hinzugezogen.

In den Simulationsstudien in Kapitel 5.5 hat sich bereits gezeigt, dass die Schätzung der Erwartungswertvektoren und Kovarianzmatrizen durch empirische Mittelwerte und Kovarianzmatrizen den übrigen Methoden bei den betrachteten Datensituationen unterlegen war. Daher wird zunächst eine Modellselektion für das Kovarianzregressionsmodell durchgeführt, wie folgt:

1. Ausgehend von einem maximalen Modell für jede Zielgröße, das Haupteffekte, Wechselwirkungen und quadratische Effekte enthält, wird zunächst eine Kombination aus Vorwärts- und Rückwärtsselektion für jede Zielgröße Y_i durchgeführt, $i = 1, \dots, p$. Dies geschieht mithilfe von univariaten generalisierten linearen Modellen auf Basis der AIC-Werte oder eines ähnlichen Kriteriums. Damit erhält man ein geschätztes Modell für den Erwartungswert jeder einzelnen Zielgröße $\hat{E}(Y_i) = g_i(f_i(x)\hat{\beta}_i)$ und für den Erwartungswertvektor $\hat{E}(Y) = g(Z_1(x)\hat{\beta})$ mit

der Matrix aus Gleichung (5.9).

2. Im nächsten Schritt wird das Modell für die Kovarianzmatrix ausgehend von $\widehat{E}(Y) = g(Z_1(x)\widehat{\beta})$ aus Schritt 1 bestimmt. Dazu wird die Matrix $Z_2(x)$ (siehe (5.30) bestimmt, wobei die Designvektoren $f_j(x) \in \{1, [x]_1, \dots, [x]_k\}$ aus allen möglichen Kombinationen von Kovariablen bestehen, $j = 1, \dots, p$. Der Interzept ist dabei immer enthalten. Die Anzahl aller Kombinationen sei $n_k \in \mathbb{N}$. Das Kovarianzregressionsmodell wird nun für alle n_k möglichen Kombinationen geschätzt und die zugehörigen AIC-Werte AIC_l für $l = 1, \dots, n_k$ berechnet.
3. Das Modell mit dem kleinsten AIC-Wert wird nun ausgewählt.

Das ausgewählte Modell lautet

$$\begin{aligned} \hat{\mu}(x_1, \dots, x_7) &= \begin{pmatrix} 720.782 + 5.251 \cdot x_1 - 3.642 \cdot x_2 + 1.989 \cdot x_3 + 1.547 \cdot x_4 + 1.778 \cdot x_5 + 1.472 \cdot x_7 \\ 0.967 + 0.332 \cdot x_1 - 0.170 \cdot x_2 + 0.107 \cdot x_3 + 0.086 \cdot x_5 + 0.032 \cdot x_7 \end{pmatrix} \\ \hat{\Sigma}(x_1, \dots, x_7) &= \begin{pmatrix} 0.184 & 0.022 \\ 0.022 & 0.0044 \end{pmatrix} \\ &+ \begin{pmatrix} -0.883 + 1.246 \cdot x_1 - 1.955 \cdot x_5 \\ 0.048 - 0.046 \cdot x_2 \end{pmatrix} \begin{pmatrix} -0.883 + 1.246 \cdot x_1 - 1.955 \cdot x_5 \\ 0.048 - 0.046 \cdot x_2 \end{pmatrix}^T. \end{aligned} \quad (8.1)$$

Abbildung 8.1 stellt die gemessenen gegen die vorhergesagten Werte für das Gewicht und die Balance dar. Bei perfekter Vorhersage sollten die Punkte auf der Winkelhalbierenden liegen. Sowohl für das Gewicht als auch für die Balance liegen die Werte entlang der Winkelhalbierenden, streuen jedoch deutlich stärker für Y_1 , was jedoch durch die Modellierung der Kovarianzmatrix in Abhängigkeit von Kovariablen berücksichtigt wird. Insgesamt ist die Vorhersage zufriedenstellend. Zusätzlich werden standardisierte Residuen der Art $e_s(x) = \Sigma(x)^{-\frac{1}{2}}(\mu(x) - y(x))$ betrachtet. Diese Residuen erfüllen die Eigenschaft $e_s(x) \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right)$. Die Einträge der geschätzten Residuen $\hat{e}_s(x)$ jeweils für das Gewicht und die Balance sind in Abbildung 8.2 gegen die vorhergesagten Werte abgetragen und in Abbildung 8.3 im zeitlichen Verlauf dargestellt. Die Residuen schwanken in beiden Grafiken scheinbar zufällig um Null und es sind keine auffälligen

Strukturen zu sehen. Daher wird das Modell bei der Optimierung im Folgenden verwendet.

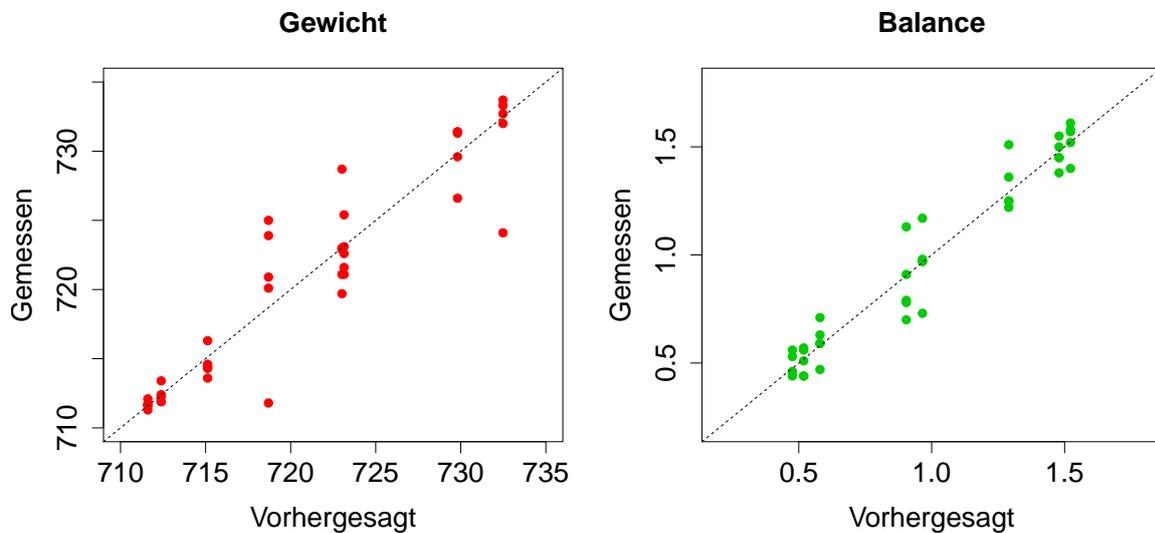


Abbildung 8.1: Vorhergesagte gegen gemessene Werte

Das ausgewählte Modell ist unabhängig von x_6 und daher wird x_6 für die weiteren Berechnungen außer Acht gelassen. Die Minimierung der Risikofunktion erfolgt unter Verwendung der Nebenbedingungen $-1 \leq x_i \leq 1$ für $i = 1, 2, 3, 4, 5, 7$. Für die Optimierung wird zunächst ein Joint Optimization Plot für unterschiedliche Folgen von Gewichtsmatrizen $W_i = \exp(a_j W_0)$ basierend auf zehn äquidistanten Werten a_1, \dots, a_{10} mit $a_1 = -5$ und $a_{10} = 5$ und

$$W_0 = \begin{pmatrix} 2 & 0.1 \\ 0.1 & 1 \end{pmatrix},$$

erzeugt. Es wird zunächst nach konstanten Bereichen in beiden Abbildungen gesucht, um anschließend den Bereich des Streckungsvektors auf den nichtkonstanten Bereich einzuschränken und dort mehr als zehn Gewichtsmatrizen zu erzeugen. In diesem Fall gibt es allerdings keinen konstanten Bereich, daher wird lediglich der Joint Optimization Plot bei Verwendung der Matrix W_0 und dreißig äquidistanten Werten a_1, \dots, a_{30}

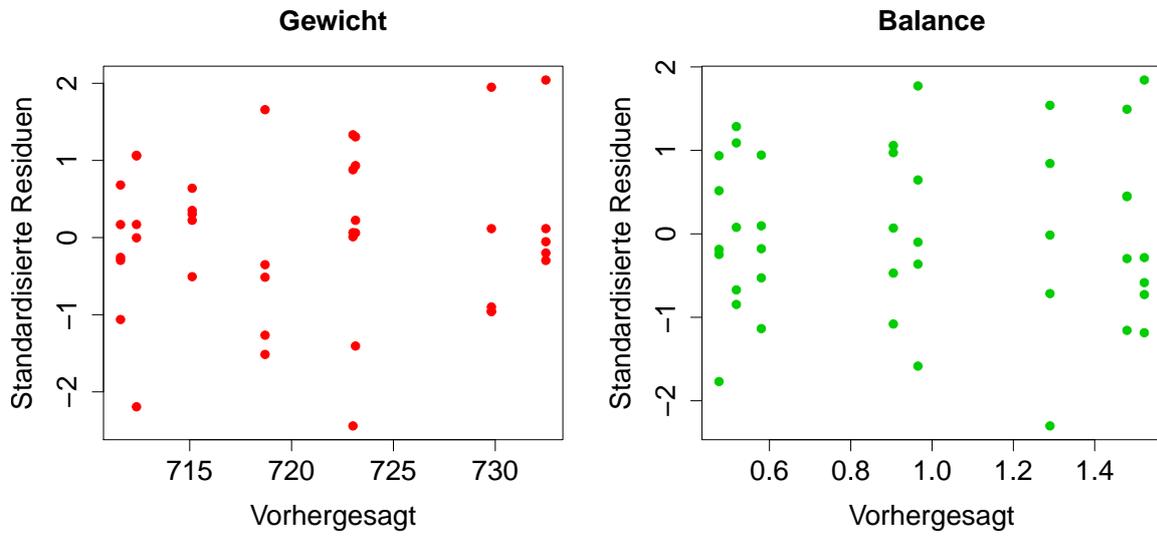


Abbildung 8.2: Vorhergesagte Werte gegen standardisierte Residuen $\hat{\epsilon}_s$

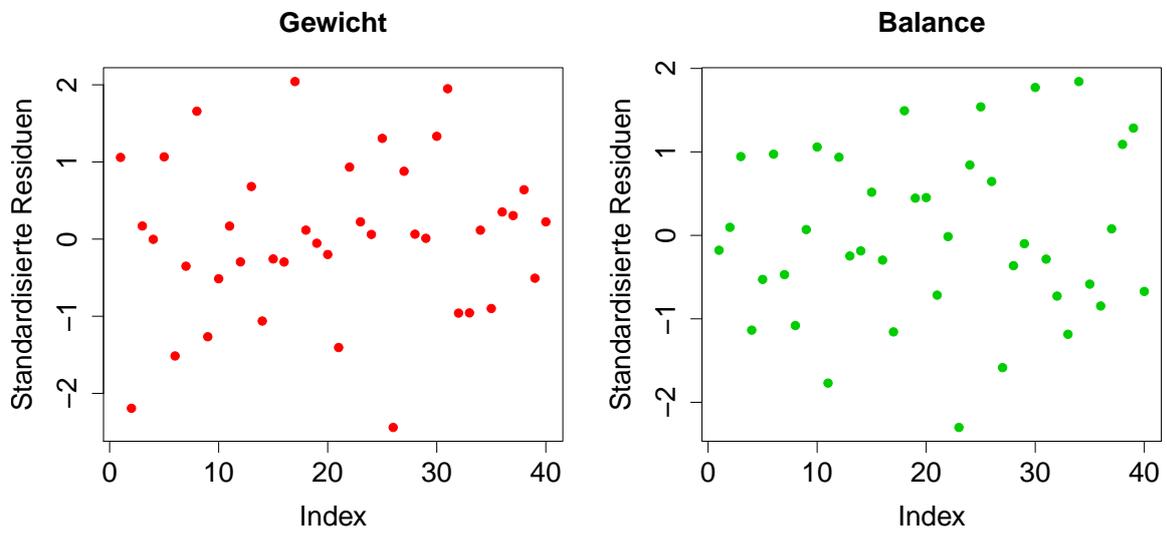


Abbildung 8.3: Standardisierte Residuen $\hat{\epsilon}_s$ im zeitlichen Verlauf

zwischen -5 und 5 in Abbildung 8.4 dargestellt. Die vorhergesagten Zielwerte haben

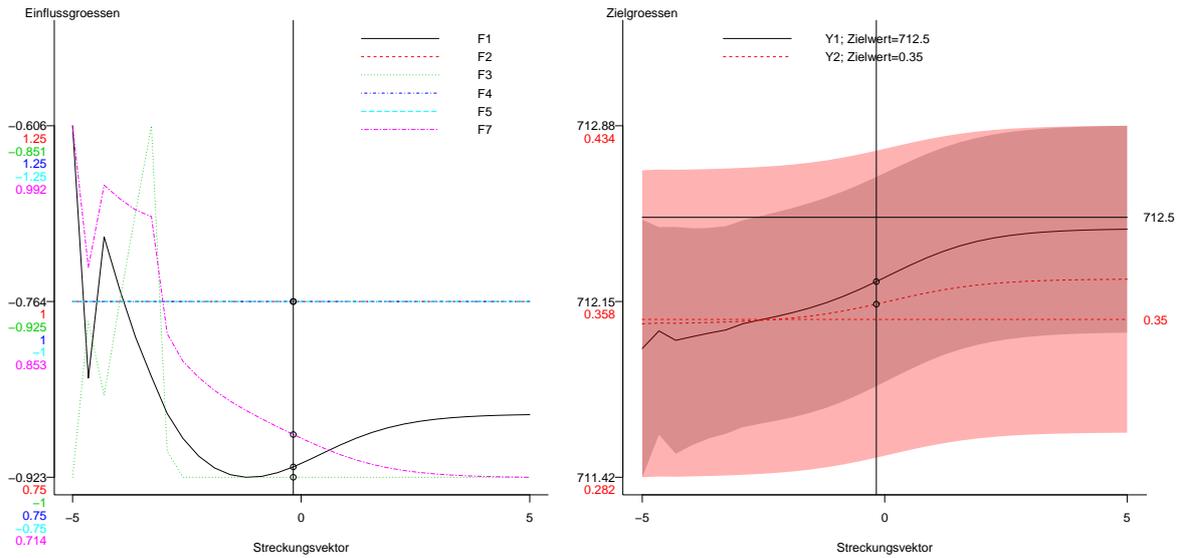


Abbildung 8.4: Joint Optimization Plot für dreißig äquidistante Werte

einen gegensätzlichen Verlauf. Auf der linken Seite der rechten Grafik sind die Werte für Y_2 nahe am Zielwert und für Y_1 maximal entfernt, während auf der rechten Seite die Werte für Y_2 maximal entfernt sind und die Werte für Y_1 nahe am Zielwert liegen. Auffällig ist zudem, dass die Werte nur geringfügig innerhalb der Bänder in der rechten Grafik von Abbildung 8.4 variieren. Daher könnten für die Anwendung durchaus mehrere Parametereinstellungen ausgewählt werden, die zu ähnlichen Vorhersagen führen. Die Kompromisslösung könnte deswegen mit dem durchführenden Ingenieur insbesondere unter Betrachtung der zugehörigen Prozessparameter erfolgen. Hier wird eine beispielhafte Kompromisslösung x_{opt} anhand des Algorithmus 7.4 ausgewählt, die auch in Abbildung 8.4 dargestellt ist.

Die ausgewählte Kompromisslösung ist in Tabelle 8.1 zusammen mit den in Salmasnia und Bashiri (2015) erhaltenen optimalen Parametereinstellungen x_{salma} . Die Lösung in Salmasnia und Bashiri (2015) liegt am Rand des experimentellen Bereichs, wie auch die in dieser Arbeit erhaltene Lösung aus Tabelle 8.1. Setzt man x_{salma} in das Modell (8.1) ein und lässt die Variable F_6 unberücksichtigt, so führt dies zu etwas schlechteren

Parametereinstellungen:							
	x_1	x_2	x_3	x_4	x_5	x_6	x_7
x_{opt}	-0.91	1.00	-1.00	1.00	-1.00		0.75
x_{salma}	-1.00	-0.928	-1.00	1.00	-1.00	1.00	0.206

Zielgrößen:				
	Y_1		Y_2	
	Vorhersage	Geschätzte Standardabweichung	Vorhersage	Geschätzte Standardabweichung
x_{opt}	712.234	0.433	0.356	0.066
x_{salma}	711.20	0.46	0.378	0.11

Geschätzte Korrelation:	
x_{opt} :	0.754
x_{salma} :	0.11

Tabelle 8.1: Kompromisslösung für das Beispiel 'Wheel Cover Experiment' und Vorhersagen basierend auf Modell (8.1)

Vorhersagen, 711.20 für Y_1 und 0.378 für Y_2 , sowie zu etwas höheren geschätzten Standardabweichungen, 0.46 für Y_1 und 0.11 für Y_2 .

8.2. Thermisches Spritzexperiment

Im folgenden wird die JOP-Methode für korrelierte Zielgrößen auf einen thermischen Spritzprozess angewandt, der bereits in Kapitel 2 beschrieben wurde. Bisherige Untersuchungen lassen vermuten, dass die Zielgrößen unter Umständen korreliert sein können (Rudak et al., 2014). Daher werden die Schichteigenschaften hier mithilfe eines multivariaten Modells modelliert, bei dem sowohl der Erwartungswertvektor als auch die Kovarianzmatrix von Kovariablen abhängen.

Beschreibung des Experiments und des Vorgehens

Die Anwendung basiert auf den Ergebnissen eines zugrundeliegenden zentral zusammengesetzten Versuchsplans für vier Einflussgrößen (Kerosin (K), Lambda (L), Spritzabstand (SOD) und Geschwindigkeit der Förderscheibe (FDV)). Der Plan enthält 30 Versuche, davon 6 Zentralpunkte und 8 Sternpunkte mit $\alpha = 2$. Die Anzahl der Experimente ist an einem Tag auf maximal 30 Versuche beschränkt. Es kann allerdings passieren, dass nicht alle 30 Versuche an einem Tag durchgeführt werden können. Daher wurde der Versuchsplan orthogonal geblockt (Myers et al., 2009) konstruiert mit $\alpha = 2$, so dass 20 Versuche am ersten Tag und 10 Versuche am zweiten Tag durchgeführt werden können. In diesem Fall konnte der komplette Versuchsplan allerdings an einem Tag durchgeführt werden. Während der Experimente werden Partikel- und Flammeigenschaften gemessen (Geschwindigkeit (Te), Temperatur (Te), Flammbreite (Wi) und Intensität (In)). Diese Werte sind unmittelbar verfügbar. Hingegen werden die aufgetragenen Schichten nachträglich durch zeitaufwändige Experimente analysiert. Diese Analysen können sich über Wochen erstrecken. Dabei werden unter anderem die Porosität (Po), Härte (Ha), Dicke (Th) und der Auftragswirkungsgrad (Dr) gemessen. Die Einfluss- und Zielgrößen sind in Tabelle 2.2 in Kapitel 2 zusammengefasst.

In bisher durchgeführten Experimenten konnte beobachtet werden, dass die Partikeleigenschaften einen messbaren Einfluss auf die Schichteigenschaften haben und Informationen über tagesabhängige Schwankungen enthalten (Rudak et al., 2012). Dabei wurden verschiedene Vorhersagestrategien miteinander verglichen, die in Abbildung 8.5 zusammengefasst sind.

Die direkte Strategie hat den Nachteil, dass tagesabhängige Schwankungen nicht erfasst werden und daher kann dieses Modell an unterschiedlichen Tagen bei gleichen Einstellungen unterschiedliche Ergebnisse liefern. Die indirekte Strategie wurde lediglich als Vergleich hinzugenommen, da die Einflussgrößen nicht kontrollierbar sondern lediglich im Prozess messbar sind. Dasselbe gilt auch für die hybride Strategie. Es hat sich heraus-

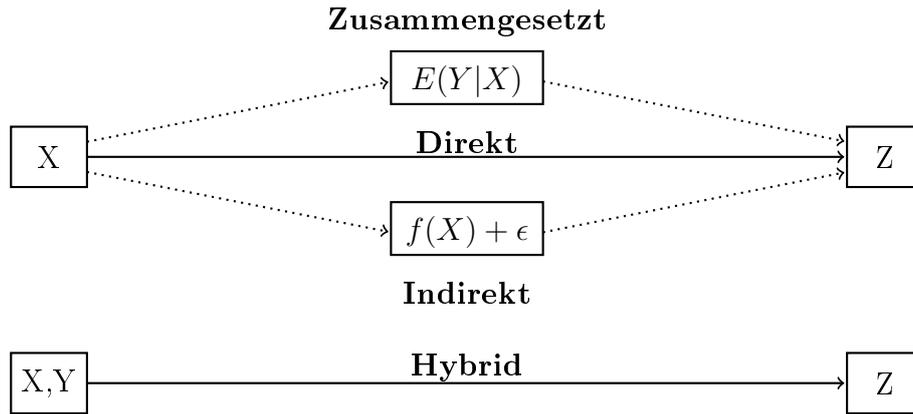


Abbildung 8.5: Vorhersagestrategien

gestellt, dass die zusammengesetzte Strategie gute Ergebnisse liefert. Dieses Vorgehen ist zudem für die Praxis sehr nützlich, da das zusammengesetzte Modell von den Einflussgrößen x abhängt und somit auch für die Optimierung von Nutzen ist. Daher wird der thermische Spritzprozess im Folgenden zweistufig mithilfe des zusammengesetzten Modells modelliert.

Rehage et al. (2012a) haben des Weiteren gezeigt, dass die Modelle der Partikeleigenschaften an unterschiedlichen Tagen durch eine einfache additive Anpassung, basierend auf wenigen Initialversuchen, angepasst werden können. Somit kann das zusammengesetzte Modell an unterschiedlichen Tagen auf Basis weniger Versuche am selben Tag adaptiert werden. Zusammenfassend wird wie folgt vorgegangen.

1. Bestimme univariate Modelle für die Partikeleigenschaften durch eine Kombination von Vorwärts- und Rückwärtsselektion auf Basis des AIC-Kriteriums und auf Basis des zentral zusammengesetzten Versuchsplans. Erhalte damit geschätzte multivariate Partikeleigenschaften $\hat{Y}_1, \dots, \hat{Y}_n \in \mathbb{R}^4$.
2. Bestimme ein multivariates Modell auf Basis der vorhergesagten Partikeleigenschaften $\hat{Y}_1, \dots, \hat{Y}_n$ entsprechend der Strategie in Abschnitt 8.1.
3. Führe eine tagesabhängige Modelladaption auf Basis der Initialversuche durch. Für

ein geschätztes Modell $\hat{E}(Y_i|x) = g(x^T \hat{\beta})$ wird lediglich ein zusätzlicher additiver Tageseffekt σ im Modell $g(x^T \hat{\beta} + \sigma)$ auf Basis der Initialversuche geschätzt. Das geschätzte Modell lautet dann $g(x^T \hat{\beta} + \hat{\sigma})$.

4. Bestimme optimale Versuchseinstellungen mithilfe der JOP-Methode.

Der zentral zusammengesetzte Versuchsplan mit Ergebnissen ist zu finden in Tabelle A.4 im Anhang und der verwendete teilfaktorielles Initialplan ist in Tabelle A.5 im Anhang abgebildet.

Prozessoptimierung mithilfe der JOP-Methode

Bezeichne mit $x = (L, K, D, FDV)$ den Parametervektor. Die adaptierten Modelle mit Tageseffekt für die Flamm- und Partikeleigenschaften ergeben sich zu:

$$\begin{aligned} \widehat{Ve}(x) &= 722.015 + 38.286 \cdot K + 10.191 \cdot L - 10.091 \cdot D - 5.332 \cdot FDV \\ &\quad - 3.829K^2 - 1.751D^2 - 4.713D \cdot FDV + 3.209K \cdot FDV + 32.77 \end{aligned}$$

$$\begin{aligned} \widehat{Te}(x) &= 1622.090 + 30.221 \cdot K - 17.531 \cdot D - 5.536 \cdot K^2 - 5.313 \cdot L \\ &\quad + 3.306FDV + 4.034D \cdot L - 3.978 \cdot D \cdot FDV + 107.9 \end{aligned}$$

$$\widehat{In}(x) = 26.437 + 5.564 \cdot FDV - 2.873 \cdot D + 1.320 \cdot K - 2.071 \cdot FDV \cdot D + 1.058$$

$$\begin{aligned} \widehat{Wi}(x) &= 9.6053 + 0.5404 \cdot FDV - 0.2608 \cdot L + 0.1944 \cdot D \\ &\quad + 0.1809 \cdot K - 0.2938 \cdot FDV \cdot D - 0.2084 \cdot FDV \cdot L + 0.2063 \cdot L \cdot D - 1.812 \end{aligned}$$

Abbildung 8.6 enthält die durch die adaptierten Partikelmodelle vorhergesagten gegen die gemessenen Werte für den Initialplan. Die Werte liegen wie gewünscht entlang der Winkelhalbierenden. Als nächstes werden die Zielgrößen Porosität und Dicke betrachtet. Es wurde zunächst eine Modellselektion durchgeführt und das resultierende zusammengesetzte Modell lautet für den Erwartungswertvektor

$$\begin{aligned} \hat{\mu}(\widehat{Ve}(x), \widehat{Te}(x), \widehat{In}(x), \widehat{Wi}(x)) &= \\ \left(\begin{array}{l} \exp(2.117 - 0.629 \cdot \widehat{Ve}(x) - 0.138 \cdot \widehat{In}(x) - 0.066 \cdot \widehat{Ve}(x)^2) \\ \exp(6.239 + 0.312 \cdot \widehat{In}(x) - 0.181 \cdot \widehat{Te}(x)^2 + 0.089 \cdot \widehat{Wi}(x) + 0.0153 \cdot \widehat{Ve}(x) \\ - 0.040 \cdot \widehat{In}(x)^2 + 0.213 \cdot \widehat{Wi}(x) \cdot \widehat{Ve}(x) \end{array} \right) \end{aligned}$$

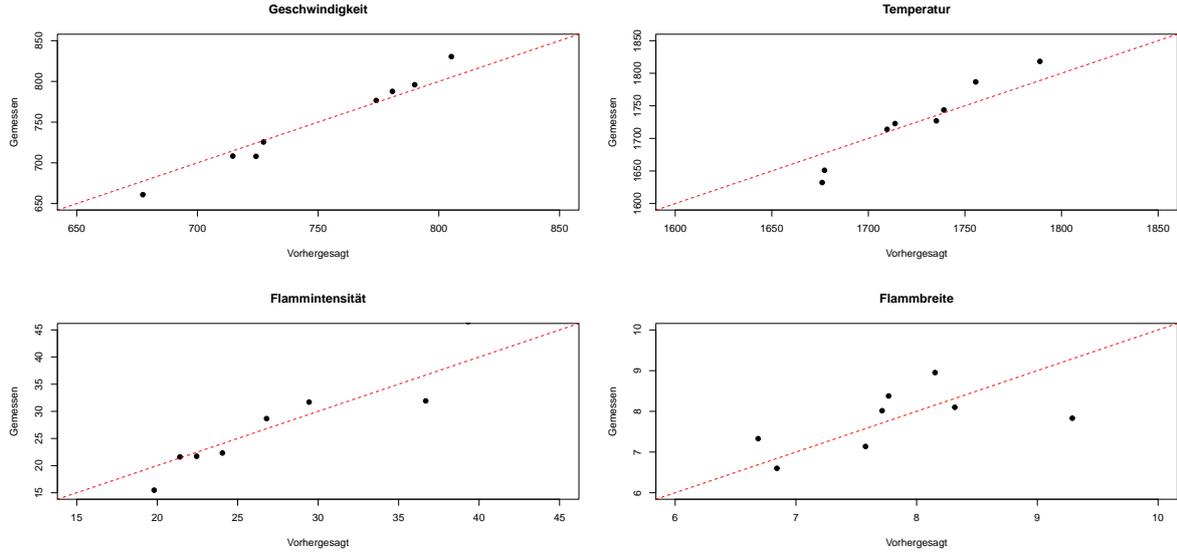


Abbildung 8.6: Vorhergesagte gegen gemessene Werte

und für die Kovarianzmatrix

$$\widehat{\Sigma}(\widehat{V}e(x), \widehat{T}e(x), \widehat{I}n(x), \widehat{W}i(x)) = \begin{pmatrix} 0.534 & 9.688 \\ 9.688 & 175.808 \end{pmatrix} + \begin{pmatrix} -4.642 + 0.158 \cdot \widehat{V}e(x) - 4.870 \cdot \widehat{W}i(x) + 0.553 \cdot \widehat{I}n(x) \\ 19.844 + 0.972 \cdot \widehat{I}n(x) \end{pmatrix} \cdot \begin{pmatrix} -4.642 + 0.158 \cdot \widehat{V}e(x) - 4.870 \cdot \widehat{W}i(x) + 0.553 \cdot \widehat{I}n(x) \\ 19.844 + 0.972 \cdot \widehat{I}n(x) \end{pmatrix}^T.$$

In Abbildung 8.7 sind die gemessenen gegen die vorhergesagten Werte für die Dicke und Porosität dargestellt. Wie im vorherigen Abschnitt 8.1 erwähnt, sollten die Werte möglichst nahe der Winkelhalbierenden liegen. Dies gelingt für die Dicke sehr zufriedenstellend und für die Porosität liegen die Werte zwar entlang der Winkelhalbierenden, streuen jedoch deutlich stärker. Es wurde bereits beobachtet, dass die Messung der Porosität einen großen Einfluss auf die Messergebnisse hat und daher möglicherweise Schwierigkeiten bei der Modellierung mit sich bringt (Hussong et al., 2012). Des Weiteren werden wie im vorherigen Abschnitt 8.1 standardisierte Residuen der Art $e_s(x) = \Sigma(x)^{-\frac{1}{2}}(\mu(x) - y(x))$ betrachtet. Die Einträge der geschätzten Residuen $\hat{e}_s(x)$

jeweils für die Dicke und Porosität sind in Abbildung 8.8 gegen die vorhergesagten Werte abgetragen. Abbildung 8.9 stellt die Residuen im zeitlichen Verlauf dar. Die Residuen schwanken in beiden Grafiken scheinbar zufällig um Null und es sind keine besonderen Auffälligkeiten zu erkennen. Daher fließt das Modell in die nachfolgende Optimierung ein.

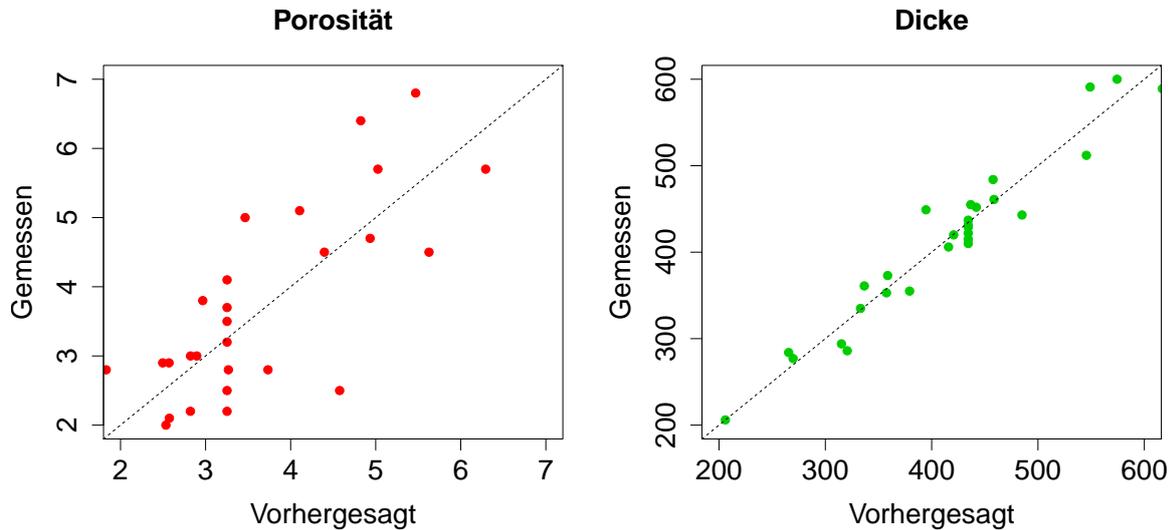


Abbildung 8.7: Vorhergesagte gegen gemessene Werte

Die Porosität ist im Allgemeinen möglichst gering zu halten und daher wird als Zielwert 0 gesetzt. Für die Dicke wird nach Absprache mit dem Ingenieur als Zielwert $300 \mu\text{m}$ gesetzt. Damit lautet der Zielwertvektor

$$\tau = \begin{pmatrix} 0 & 300 \end{pmatrix}^T.$$

Die Optimierung wird innerhalb der experimentellen Region durchgeführt, das heißt mit den Nebenbedingungen $-2 \leq x_i \leq 2$ für $i = 1, \dots, 4$. Es werden Gewichtsmatrizen auf Basis der Matrix

$$W_0 = \begin{pmatrix} 2 & 0.1 \\ 0.1 & 1 \end{pmatrix},$$

und zehn äquidistanten Werten a_1, \dots, a_{10} mit $a_1 = -5$ und $a_{10} = 5$ berechnet. Der zugehörige Joint Optimization Plot ist in Abbildung 8.10 zu sehen. Auffällig ist, dass

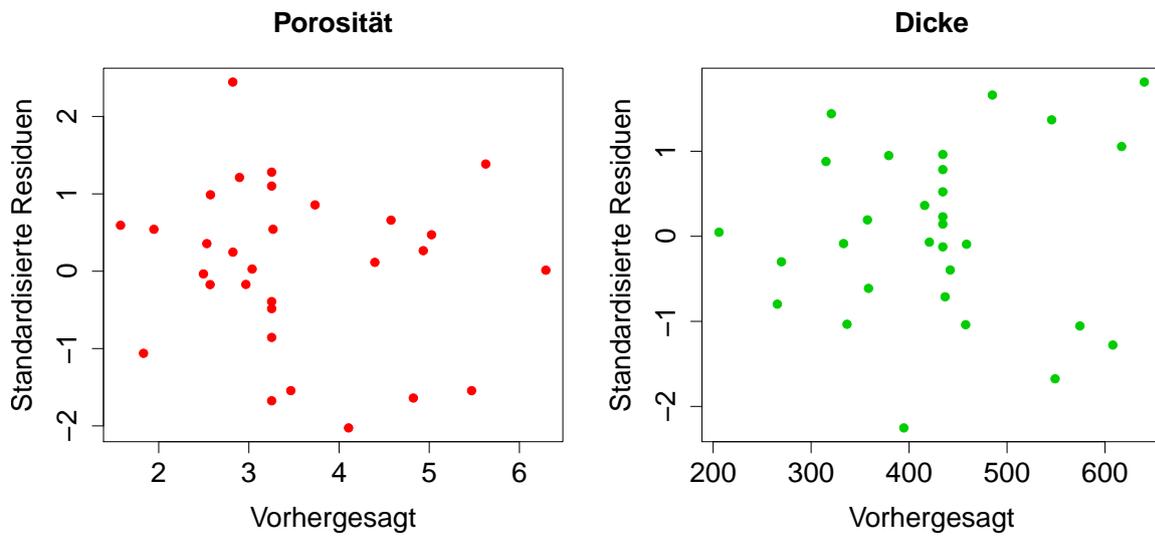


Abbildung 8.8: Vorhergesagte Werte gegen standardisierte Residuen $\hat{\epsilon}_s$

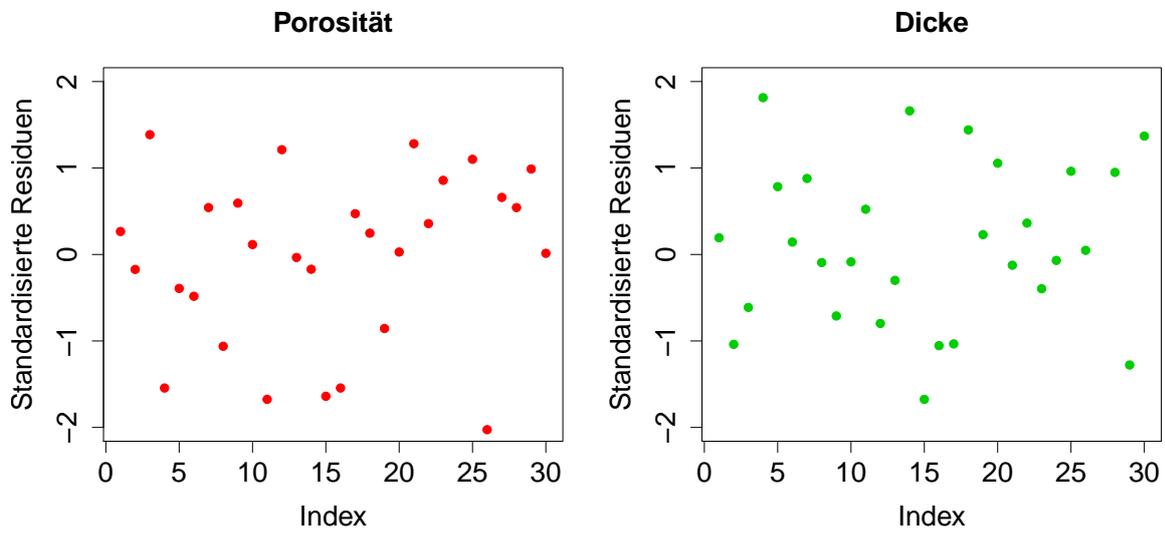


Abbildung 8.9: Standardisierte Residuen $\hat{\epsilon}_s$ im zeitlichen Verlauf

die Werte auf der linken Seite beider Grafiken konstant sind. Daher wird der Bereich des Streckungsvektors eingeschränkt und Gewichtsmatrizen mit W_0 und 30 äquidistanten Werten a_1, \dots, a_{30} mit $a_1 = -2$ und $a_{30} = 5$ berechnet. Der zugehörige Joint Optimization Plot ist in Abbildung 8.11 dargestellt. Die geschätzte Standardabweichung

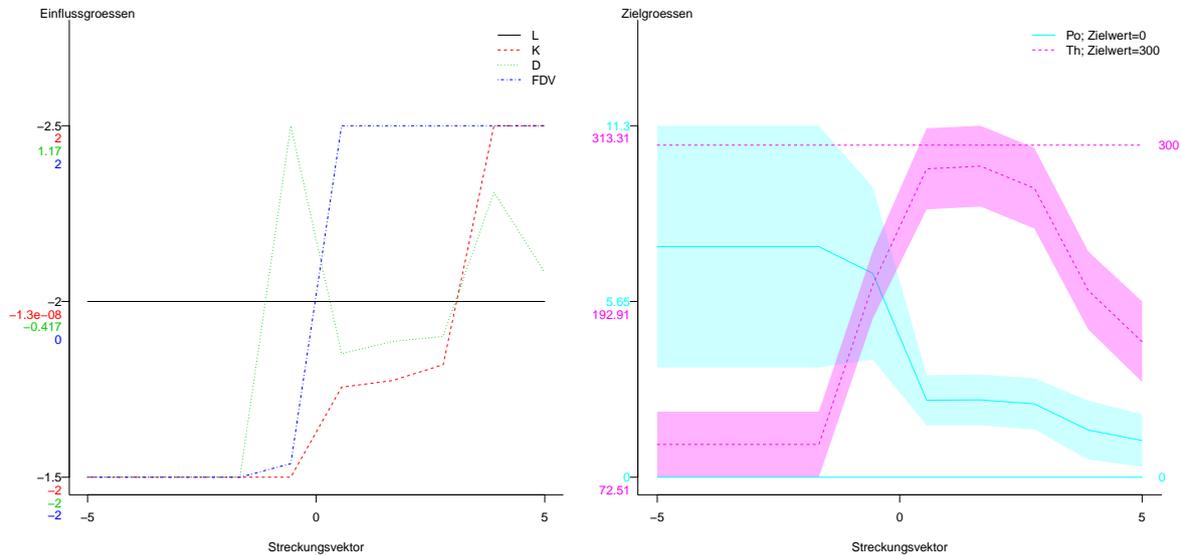


Abbildung 8.10: Joint Optimization Plot für zehn äquidistante Werte

ist für die Porosität auf der linken Seite deutlich größer als auf der rechten. Zur rechten Seite hin etwa fällt die Standardabweichung für die Porosität deutlich. Dagegen ist die geschätzte Standardabweichung für die Dicke auf der linken Seite der rechten Grafik etwas kleiner als auf der rechten. In der Mitte der rechten Grafik nähern sich die Lösungen simultan ihrem Zielwert. Danach entfernt sich erneut die Vorhersage der Zielgröße Dicke. Ein geeigneter Kompromiss wäre hier nach Absprache mit dem Ingenieur möglicherweise in der Mitte zu finden. Dazu passt auch die anhand von Algorithmus 7.4 ausgesuchte Kompromisslösung, die zudem in Abbildung 8.11 eingezeichnet ist. Die ausgewählte Kompromisslösung mit geschätzten Zielgrößen, Standardabweichungen und geschätzter Korrelation ist in Tabelle 8.2 zusammengefasst. Im nächsten Schritt müssten die Vorhersagen in echten Experimenten verifiziert werden.

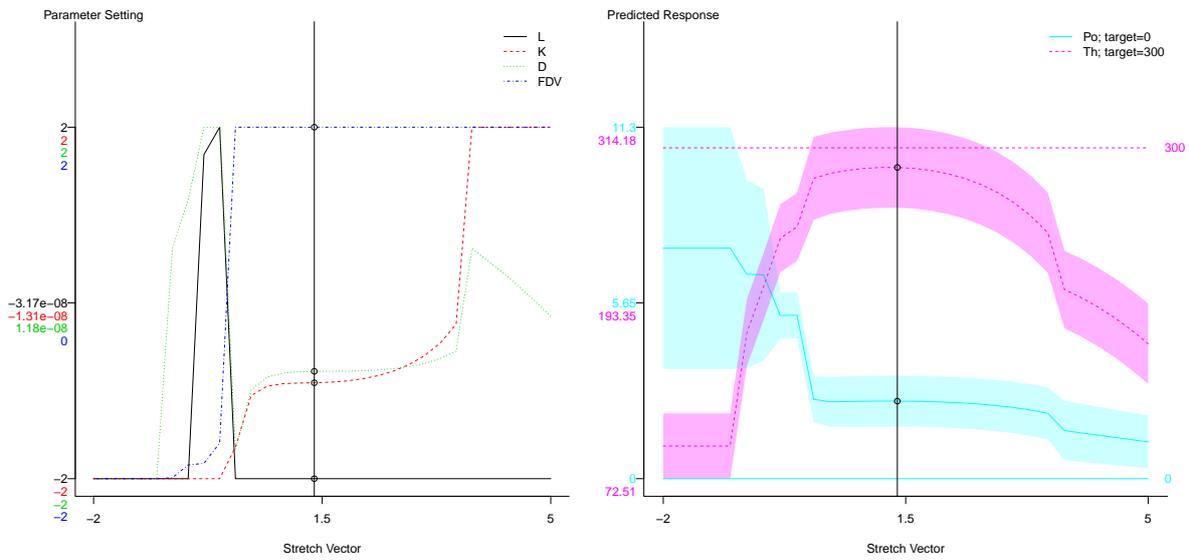


Abbildung 8.11: Joint Optimization Plot für dreißig äquidistante Werte

Parametereinstellungen

L	K	D	FDV
-2.00	-0.91	-0.78	2.00

Zielgrößen

Porosität		Dicke	
	Geschätzte		Geschätzte
Vorhersage	Standardabweichung	Vorhersage	Standardabweichung
2.49	0.81	286.50	27.68

Geschätzte Korrelation: 0.82

Tabelle 8.2: Kompromisslösung für thermischen Spritzprozess

9. Zusammenfassung und Ausblick

In dieser Arbeit wurden zunächst verschiedene Methoden zur Optimierung multipler Zielgrößen vorgestellt. Dazu zählt auch die JOP-Methode, bei der zunächst die einzelnen Zielgrößen mit doppelt generalisierten linearen Modellen abgebildet werden und anschließend eine Risikofunktion für eine ganze Reihe von Kostenmatrizen minimiert wird. Die klassische JOP-Methode ist zunächst für unkorrelierte Zielgrößen geeignet. Es wurde im Verlauf der Arbeit eine Wahl einer diagonalen Kostenmatrix vorgeschlagen, die sich aus dem Produkt einer Standardisierungsmatrix und einer Gewichtsmatrix zusammensetzt. Die Standardisierungsmatrix sorgt dafür, dass die Risikofunktion invariant gegenüber linearen Transformationen bleibt (wie etwa verschiedene Einheiten, wie cm oder mm). Zudem wurde die Pareto-Optimalität für die JOP-Methode nachgewiesen, wie bereits in Erdbrügge et al. (2011) publiziert. Des Weiteren wurde eine grafische Interpretation der JOP-Methode vorgestellt sowie die Implementierung der JOP-Methode in Form des R-Pakets JOP präsentiert (Kuhnt und Rudak, 2013b).

Im weiteren Verlauf wurde die JOP-Methode auf korrelierte Zielgrößen erweitert. Dies wurde durch die Weiterentwicklung multivariater Modelle möglich, bei denen sowohl der Erwartungswertvektor als auch die Kovarianzmatrix von Kovariablen abhängen darf. Das Kovarianzregressionsmodell (Hoff und Niu, 2012) wurde mithilfe einer alternativen Parametrisierung dargestellt. Zudem wurde das Modell durch Verwendung einer nichtlinearen Linkfunktion erweitert. Es wurde ein Algorithmus auf Basis des EM-Algorithmus zur ML-Schätzung des Modells hergeleitet. Zudem wurde das Matrix-Logarithmus-Modell (Chiu et al., 1996) durch Verwendung einer nichtlinearen Linkfunktion erweitert. Auch bei diesem Modell wurde ein zweistufiger Algorithmus zur ML-Schätzung basierend auf dem Newton-Verfahren mit Schrittweitsuche hergeleitet. Des Weiteren wurden Grenzwerte für einige sich ergebende Ausdrücke bei der numerischen Umsetzung berechnet, um numerische Probleme bei der Implementierung zu verhindern. Das nichtparametrische Modell (Yin et al., 2010) wurde hier für einen k -dimensionalen Kovariablenvektor eingesetzt. Angelehnt an das nichtparametrische Modell wurde eine semiparametrische

Heuristik als Vergleich hinzugenommen. Somit kann aus einer breiten Sammlung von Modellen das geeignete multivariate Modell ausgewählt werden. In Simulationsstudien wurden die Schätzung der vorgestellten Modelle analysiert und verglichen. Als Referenz wurde hier ein einfaches Vorgehen, bei dem der empirische Mittelwertvektor und die empirische Kovarianzmatrix für jede Einstellung berechnet wird, hinzugezogen. Die Simulationsstudie ergab zwar keinen klaren Favoriten, allerdings wurde das einfache Vorgehen in jeder betrachteten Situation von mindestens einem der übrigen Modelle übertroffen. Das Kovarianzregressionsmodell stellt insbesondere wegen seiner einfachen und daher leichter interpretierbaren Struktur einen guten Kandidaten für praktische Anwendungen dar.

Im weiteren Verlauf wurde eine mögliche Wahl von nichtdiagonalen Kostenmatrizen vorgeschlagen. Des Weiteren wurde ein Kriterium zur Bestimmung einer Kompromisslösung präsentiert und die Pareto-Optimalität auch im Falle von korrelierten Zielgrößen und nichtdiagonalen Kostenmatrizen nachgewiesen. Daraufhin wurde die erweiterte JOP-Methode auf einen Datensatz aus der Literatur sowie erfolgreich auf einen Datensatz aus dem thermischen Spritzen angewandt.

Sowohl auf der Modellierungs- als auch auf der Prozessoptimierungsebene ergeben sich weitere Forschungsfragen. Zunächst können Schätzer für die vorgestellten parametrischen Modelle beispielsweise unter Annahme einer alternativen multivariaten Verteilung, wie etwa der t-Verteilung, hergeleitet werden, um die Annahme einer multivariaten Normalverteilung abzuschwächen. Interessant ist hier etwa eine kürzlich in Bonat und Jørgensen (2015) vorgeschlagene multivariate Erweiterung des GLM. Bonat und Jørgensen (2015) schlagen ein multivariates GLM für den Erwartungswertvektor und die Kovarianzmatrix ohne Angabe einer konkreten Verteilungsannahme vor. Das Modell enthält eine Linkfunktion sowohl für den Erwartungswertvektor als auch für die Kovarianzmatrix. Die Modellparameter werden lediglich auf Basis der ersten beiden Momente geschätzt. Unter Umständen ist es möglich, das Kovarianzregressionsmodell und das Matrix-Logarithmus-Modell auf alternative Verteilungen ähnlich wie bei Bonat und Jørgensen (2015) zu

verallgemeinern und einen ähnlichen Algorithmus zur Schätzung der Modellparameter herzuleiten.

Zudem kann die Vorhersagegenauigkeit der Modelle durch eine verbesserte gezielte Modellselektion sicher verbessert werden. Dazu können unterschiedliche Modellselektionsstrategien, basierend auf Informationskriterien oder etwa LASSO, hergeleitet und verglichen werden. Das nichtparametrische Modell kann unter Umständen durch eine andere Wahl der Kernfunktion oder eine bessere Wahl der Bandbreite verbessert werden.

Hinsichtlich der Prozessoptimierung kann für die JOP-Methode eine alternative Wahl der Verlustfunktion in Betracht gezogen werden und mit der JOP-Methode basierend auf der quadratischen Verlustfunktion verglichen werden. Des Weiteren ist die Wahl alternativer Kostenmatrizen von Interesse. In Anwendungen kommt es vielfach vor, dass konstante Bereiche im Joint Optimization Plot zu erkennen sind. Eine automatische Erkennung solcher Bereiche und adaptive Anpassung des Streckungsvektors wäre für den Anwender hilfreich.

Literatur

- J. Antony. Multi-response optimization in industrial experiments using Taguchi's quality loss function and principal component analysis. *Quality and Reliability Engineering International*, 16:3–8, 2000.
- M.K. Ardakani und S.S. Wulff. Overview of optimization formulations for multiresponse surface problems. *Quality and Reliability Engineering International*, 29(1):3–16, 2013.
- F.-W. Bach, A. Laarmann, und T. Wenz. *Modern Surface Technology*. Wiley VCH, 1st edition, 2006.
- R. Bellman. *Introduction to Matrix Analysis*. McGraw-Hill, New York, 1970.
- W.H. Bonat und B. Jørgensen. Multivariate covariance generalized linear models. Submitted, 2015.
- M. Borowski, N. Rudak, B. Hussong, D. Wied, S. Kuhnt, und W. Tillmann. On- and offline detection of structural breaks in thermal spraying processes. *Journal of Applied Statistics*, 41(5):1073–1090, 2014.
- J. Chandra. *Statistical Quality Control*. CRC Press, 2001.
- C.-H. Chiao und M. Hamada. Analyzing experiments with correlated multiple responses. *Journal of Quality Technology*, 33(4):451–465, 2001.
- T.Y.M. Chiu, T. Leonard, und K.-W. Tsui. The matrix-logarithmic covariance model. *Journal of the American Statistical Association*, 91(433):198–210, 1996.
- N.R. Costa, J. Lourenco, und Z.L. Pereira. Desirability function approach: A review and performance evaluation in adverse conditions. *Chemometrics and Intelligent Laboratory Systems*, 107:234–244, 2011.
- A.P. Dempster, N.M. Laird, und D.B. Rubin. Maximum-likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, 39(1):1–38, 1977.

- G. Derringer und R. Suich. Simultaneous optimization of several response variables. *Journal of Quality Technology*, 12(4):214–219, 1980.
- Peter K Dunn. *tweedie: Tweedie exponential family models*, 2014. R package version 2.2.1.
- M. Erdbrügge und S. Kuhnt. Strategies for multi-response parameter design using loss functions and joint optimization plots. SFB 475, technical report 8/2002, Universität Dortmund, 2002.
- M. Erdbrügge, S. Kuhnt, und N. Rudak. Joint optimization of independent multiple responses. *Quality and Reliability Engineering International*, 27:689–704, 2011.
- L. Fahrmeir und H. Kaufmann. Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. *Annals of Statistics*, 13:342–368, 1985.
- L. Fahrmeir und G. Tutz. *Multivariate Statistical Modeling based on Generalized Linear Models (Second edition)*. Springer-Verlag New York, 2001.
- F. Gao, X. Huang, R. Liu, und Q. Yang. Optimization of plasma spray process using statistical methods. *Journal of Thermal Spray Technology*, 21(1):176–186, 2012.
- W.W. Hager. Updating the inverse of a matrix. *SIAM Review*, 31(2):221–239, 1989.
- M. Hamada und N. Balakrishnan. Analyzing unreplicated factorial experiments: A review with some proposals. *Statistica Sinica*, 8:1–41, 1998.
- S. Hasan und J. Stokes. Design of experiment analysis of the sulzer metco dj high velocity oxy-fuel coating of hydroxyapatite for orthopedic applications. *Journal of Thermal Spray Technology*, 20(1-2):186–194, 2011.
- C. Hillermeier. *Nonlinear Multiobjective Optimization*. Birkhäuser Verlag, Basel, 2001.
- D.V. Hinkley, N. Reid, und E.J. Snell. *Statistical Theory and Modelling*. Chapman & Hall, 1991.

- P.D. Hoff und X. Niu. A covariance regression model. *Statistica Sinica*, 22:729–753, 2012.
- J.Z. Huang, N. Liu, M. Pourahmadi, und L. Liu. Covariance matrix selection and estimation via penalised normal likelihood. *Biometrika*, 93(1):85–98, 2006.
- B. Hussong, W. Tillmann, S. Kuhnt, N. Rudak, und A. Rehage. Vergleich unterschiedlicher Strategien zur Prozessmodellierung beim HVOF thermischen Spritzen. In *Ta- gungsband zum 15. Werkstofftechnischen Kolloquium*, pages 68–73, Chemnitz, 2012.
- B. Jorgensen. Exponential dispersion models. *Journal of the Royal Statistical Society B*, 49:127–162, 1987.
- H. Kaßner, E. Höck, T. Gartner, O. Schmäscke, und S. Deng. High efficient thermal spraying processes and strategies for aviation industry. In *Conference Proceedings der ITSC 2011: DVS-Berichte Volume 276*, 2011.
- A.I. Khuri und M. Conlon. Simultaneous optimization of multiple responses represented by polynomial regression functions. *Technometrics*, 23(4):363–375, 1981.
- Y.-H. Ko, K.-J. Kim, und C.-H. Jun. A new loss function-based method for multiresponse optimization. *Journal of Quality Technology*, 37(1):50–59, 2005.
- K. Königsberger. *Analysis 1*. Springer, 2004.
- S. Kuhnt und M. Erdbrügge. A strategy of robust parameter design for multiple responses. *Statistical Modelling*, 9:249–264, 2004.
- S. Kuhnt und N. Rudak. *JOP: Joint Optimization Plot*, 2013a. URL <http://ouch.r-forge.r-project.org>. R package version 3.6.
- S. Kuhnt und N. Rudak. Simultaneous Optimization of Multiple Responses with the R Package JOP. *Journal of Statistical Software*, 54(9), 2013b.
- T. Leonard und J.S.J. Hsu. Bayesian inference for a covariance matrix. *The Annals of Statistics*, 20(4):1669–1696, 1992.

- P. McCullagh und J.A. Nelder. *Generalized linear models (Second edition)*. London: Chapman & Hall, 1999.
- G.:J. McLachlan und T. Krishnan. *The EM algorithm and extensions*. Wiley series in probability and statistics. Wiley, Hoboken, NJ, 2. ed edition, 2008.
- S. Mukhopadhyay und A.I. Khuri. Optimization in a multivariate generalized linear model situation. *Computational Statistics and Data Analysis*, 52:4625–4634, 2008.
- R.H. Myers, D.C. Montgomery, und C.M. Anderson-Cook. *Response Surface Methodology*. Wiley, 3rd edition, 2009.
- J.A. Nelder und R.W.M. Wedderburn. Generalized linear models. *Journal of the Royal Statistical Society Series A (General)*, 135:370–384, 1972.
- G. Papageorgiou. Restricted maximum likelihood estimation of joint mean-covariance models. *The Canadian Journal of Statistics*, 40(2):225–242, 2012.
- J.J. Pignatiello. Strategies for robust multiresponse quality engineering. *IIE Transactions*, 25(3):5–15, 1993.
- J.C. Pinheiro und D.M. Bates. Unconstrained parametrizations for variance-covariance matrices. *Statistics and Computing*, 6:289–296, 1996.
- M. Pourahmadi. Joint mean-covariance models with applications to longitudinal data: Unconstrained parametrisation. *Biometrika*, 86(3):677–690, 1999.
- M. Pourahmadi. Maximum likelihood estimation of generalised linear models for multivariate normal covariance models. *Biometrika*, 87(2):425–435, 2000.
- M. Pourahmadi. Covariance estimation: The glm approach and regularization perspectives. *Statistical Science*, 26(3):369–387, 2011.
- C.S. Ramachandran, V. Balasubramanian, und P.V. Ananthapadmanabhan. Multiobjective optimization of atmospheric plasma spray process parameters to deposit yttria-

- stabilized zirconia coatings using response surface methodology. *Journal of Thermal Spray Technology*, 20(3):590–607, 2011.
- A. Rehage, N. Rudak, B. Hussong, S. Kuhnt, und W. Tillmann. Prediction of in-flight particle properties in thermal spraying with additive day-effects. SFB 823 discussion paper 06/12, TU Dortmund University, 2012a.
- A. Rehage, N. Rudak, B. Hussong, S. Kuhnt, und W. Tillmann. Einführung eines Tageseffekt-Schätzers zur Verbesserung der Vorhersage von Partikeleigenschaften in einem HVOF-Spritzstrahl. *Thermal Spray Bulletin*, 5(2), 2012b.
- R. Reinhardt, A. Hoffmann, und T. Gerlach. *Nichtlineare Optimierung - Theorie, Numerik und Experimente*. Springer Spektrum, 2013.
- N. Rudak, S. Kuhnt, B. Hussong, und W. Tillmann. On different strategies for the prediction of coating properties in a HVOF process. SFB 823 discussion paper 29/12, TU Dortmund University, 2012.
- N. Rudak, S. Kuhnt, und E. Riccomagno. Numerical algebraic fan of a design for statistical model building. SFB 823 discussion paper 04/13, TU Dortmund University, 2013.
- N. Rudak, S. Kuhnt, und E. Riccomagno. Modeling of a Thermal Spraying Process by Gaussian Chain Graphs. *Quality Technology and Quantitative Management*, 11(1): 85–98, 2014.
- N. Rudak, B. Hussong, und S. Kuhnt. Simultaneous optimization of multiple correlated responses with application to a thermal spraying process. *Quality and Reliability Engineering International*, 2015a. akzeptiert zur Veröffentlichung.
- N. Rudak, S. Kuhnt, und E. Riccomagno. Numerical algebraic fan of a design for statistical model building. *Statistica Sinica*, 2015b. doi: 10.5705/ss.2014.264.

- J. Saaedi, T.W. Coyle, H. Arabi, S. Mirdamadi, and J. Mostaghimi. Effects of HVOF process parameters on the properties of Ni-Cr coatings. *Journal of Thermal Spray Technology*, 19(3):521–530, 2010.
- A. Salmasnia und M. Bashiri. A new desirability function-based method for correlated multiple response optimization. *International Journal of Advanced Manufacturing Technology*, 76:1047–1062, 2015.
- A. Salmasnia, R.B. Kazemzadeh, und S.T.A. Niaki. An approach to optimize correlated multiple responses using principal component analysis and desirability function. *International Journal of Advanced Manufacturing Technology*, 62:835–846, 2012.
- A. Salmasnia, M. Bashiri, und M. Salehi. A robust interactive approach to optimize correlated multiple responses. *International Journal of Advanced Manufacturing Technology*, 67:1923–1935, 2013a.
- A. Salmasnia, R.B. Kazemzadeh, M. Seyyed-Esfahani, und T.H. Hejazi. Multiple response surface optimization with correlated data. *International Journal of Advanced Manufacturing Technology*, 64:841–855, 2013b.
- K. Schmidt und G. Trenkler. *Einführung in die moderne Matrix-Algebra*. Springer, 2006.
- B.W. Silverman. *Density Estimation for Statistics and Data Analysis*. Chapman&Hall, 1998.
- G.K. Smyth. Generalized linear models with varying dispersion. *Journal of the Royal Statistical Society. Series B (Methodological)*, 51(1):47–60, 1989.
- G.K. Smyth und P.K. Dunn. Evaluation of tweedie exponential dispersion model densities by fourier inversion. *Statistical Computing*, 18:73–86, 2008.
- G.K. Smyth und A.P. Verbyla. Adjusted likelihood methods for modelling dispersion in generalized linear models. *Environmetrics*, 10(6):695–709, 1999.

- G. Taguchi, E.A. Elsayed, , und T. Hsiang. *Quality Engineering in Production Systems*. McGraw-Hill, New York, 1989.
- W. Tillmann, E. Vogli, I. Baumann, G. Kopp, und C. Weihs. Desirability-based multi-criteria optimization of HVOF spray experiments to manufacture fine structured wear-resistant 75Cr3C2-25(NiCr20) coatings. *Journal of Thermal Spray Technology*, 19(1-2):392–408, 2010.
- G.G. Vining. A compromise approach to multiresponse optimization. *Journal of Quality Technology*, 30:309–313, 1998.
- L. Wasserman. *All of Statistics: A Concise Course in Statistical Inference*. Springer, New York, 2004.
- H. Witting. *Mathematische Statistik 1*. B.G. Teubner Stuttgart, 1985.
- Kris Wonggasem. *Ensuring an Efficient Turning Process by Means of Desirability Index Optimization for Correlated Quality Criteria*. Dissertation, Technische Universität Dortmund, 2012.
- F.C. Wu. Optimization of correlated multiple quality characteristics using desirability function. *Quality Engineering*, 17:119–126, 2005.
- F.C. Wu. Robust design of nonlinear multiple dynamic quality characteristics. *Computers and Industrial Engineering*, 56(4):1328–1332, 2009.
- J. Yin, Z. Geng, R. Li, und H. Wang. Nonparametric covariance model. *Statistica Sinica*, 20:469–479, 2010.

A. Appendix

A.1. Newton-Algorithmus

Im folgenden wird ein gedämpfter Newton-Algorithmus (Reinhardt et al., 2013) kurz erläutert. Sei dazu exemplarisch eine mindestens zweimal stetig differenzierbare Funktion $f : x \in \mathbb{R}^k \rightarrow f(x) \in \mathbb{R}$ gegeben. Der gedämpfte Newton-Algorithmus durchläuft folgende Schritte. Ausgehend von einem Iterationspunkt $x^{(s)} \in \mathbb{R}^k$ wird im Schritt (s+1) ein Newton-Schritt

$$d^{(s+1)} = \left(\frac{\partial^2}{\partial x \partial x^T} f(x) \Big|_{x=x^{(s)}} \right)^{-1} \frac{\partial}{\partial x} f(x) \Big|_{x=x^{(s)}} \quad (\text{A.1})$$

berechnet. Anschließend wird eine Schrittweite $t^{(s+1)} > 0$ bestimmt mit $f(x^{(s)}) > f(x^{(s)} - t^{(s+1)}d^{(s+1)})$. Der neue Iterationspunkt lautet $x^{(s+1)} = x^{(s)} - t^{(s+1)}d^{(s+1)}$. Es gibt verschiedene Möglichkeiten zur Schrittweitenbestimmung, die einen hinreichend großen Abstieg erzielen, wie etwa die Armijo-Schrittweite (Reinhardt et al., 2013). Die Armijo-Schrittweite kann wie folgt bestimmt werden.

Algorithmus A.1 (Armijo-Schrittweite) Gegeben sei $x_0 \in \mathbb{R}^n$ und eine Abstiegsrichtung $d \in \mathbb{R}^n$ in x für eine mindestens zweimal stetig differenzierbare Funktion $f : x \in \mathbb{R}^k \rightarrow f(x) \in \mathbb{R}$, d.h. $\frac{\partial}{\partial x} f(x) \Big|_{x=x_0}^T d < 0$.

1. Wähle $\psi_1, \psi_2 \in (0, 1)$. Setze $l=1$.
2. Berechne $t^{(l)} = (\psi_1)^l$.
3. Falls gilt $f(x_0 + t^{(l)}d) < f(x_0) + \psi_2 t^{(l)} \frac{\partial}{\partial x} f(x) \Big|_{x=x_0}^T d$, so gehe zu 4. Andernfalls setze $l = l + 1$ und gehe zu 2.
4. Die Armijo-Schrittweite lautet $t_{\text{Armijo}} = t^{(l)}$.

Der Newton-Algorithmus lässt sich wie folgt zusammenfassen.

Algorithmus A.2 (Newton-Algorithmus)

1. Wähle einen Startvektor $x^{(0)}$ und eine Schranke $\epsilon > 0$. Setze $s = 0$.
2. Berechne $d^{(s)} = (\frac{\partial^2}{\partial x \partial x^T} f(x)|_{x=x^{(s)}})^{-1} \frac{\partial}{\partial x} f(x)|_{x=x^{(s)}}$ und eine Schrittweite $t^{(s)} > 0$ mit $f(x^{(s)}) > f(x^{(s)} - t^{(s)}d^{(s)})$. Setze $x^{(s+1)} = x^{(s)} + t^{(s)}d^{(s)}$.
3. Falls Abbruchkriterium erfüllt, beispielsweise $\|\frac{\partial}{\partial x} f(x)|_{x=x^{(s+1)}}\| < \epsilon$, so breche ab. Andernfalls setze $s = s + 1$ und gehe zu 2.

A.2. JOP für unkorrelierte Zielgrößen: Liste der Eingabe- und Ausgabeargumente

Kategorie	Argument	Erklärung
Allgemein	<code>datax</code>	Beispielhafter Datensatz mit statistischem Versuchsplan
	<code>datay</code>	Beispielhafter Datensatz mit Zielgrößen
	<code>jopplot</code>	Erzeugung des Joint Optimization Plot, wenn <code>jopplot = TRUE</code>
Folge von Kostenmatrizen	<code>Wstart, Wend, numbW, d</code>	Parameter für Kostenmatrizen, siehe (6.5), T entspricht <code>numbW</code> , <code>Wstart</code> entspricht $\log a_1$ und <code>Wend</code> entspricht $\log a_T$ und
	<code>form.mean</code>	Liste von Formeln für Erwartungswerte jeder Zielgröße (Typ <code>formula</code>)
	<code>form.disp</code>	Liste von Formeln für Dispersion jeder Zielgröße (Typ <code>formula</code>)

Modelle für Erwartungswert und Dispersion	<code>family.mean</code>	Verteilungsfamilie für Erwartungswert
	<code>dlink</code>	Liste von Linkfunktionen für Dispersion jeder Zielgröße
	<code>mean.model</code>	Liste von Funktionen für den Erwartungswert jeder Zielgröße
	<code>var.model</code>	Liste von Funktionen für die Varianz jeder Zielgröße
Optimierung	<code>optreg</code>	<code>optreg</code> bezeichnet den Typ der Nebenbedingungen ("box" für obere und untere Grenzen und "sphere" für sphärische Nebenbedingungen)
	<code>solver</code>	"solnp" oder "gosolnp"
	<code>tau</code>	<code>tau</code> ist der Vektor der zu erreichenden Zielwerte
	<code>Domain</code>	Obere und untere Grenzen für die Kovariablen

Tabelle A.1: Eingabeargumente der Funktion JOP

Ausgabe	Erklärung
Parameters	Liste der Pareto-optimalen Kovariablen
Responses	Liste der Pareto-optimalen Zielgrößen
StandardDeviation	Liste der geschätzten Standardabweichungen
OptimalValue	Minimale Risikowerte
TargetValueJOP	Intern durch JOP verwendete Zielwerte
TargetValueUSER	Durch den Anwender übergebene Zielwerte
DGLM	Liste der angepasste doppelt generalisierte lineare Modelle
RiskminimalParameters	Risikominimale Kovariablen im Sinne von (6.10)
RiskminimalResponses	Risikominimale Zielgrößen im Sinne von (6.10)
d	siehe Tabelle A.1
d	siehe Tabelle A.1

Tabelle A.2: Ausgabeargumente der Funktion JOP

A.3. Parameter für die Simulationsstudie in Kapitel 5.5

Die Matrizen für den Kovarianztyp 1 in der Simulationsstudie in Kapitel 5.5 werden wie folgt am Beispiel von 2 Kovariablen und 3 Zielgrößen festgelegt.

$$\begin{array}{l}
 U_1(x) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad U_2(x) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad U_3(x) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \\
 U_4(x) = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad U_5(x) = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \quad U_6(x) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \\
 U_7(x) = \begin{pmatrix} [x]_1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad U_8(x) = \begin{pmatrix} 0 & [x]_1 & 0 \\ [x]_1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad U_9(x) = \begin{pmatrix} 0 & 0 & [x]_1 \\ 0 & 0 & 0 \\ [x]_1 & 0 & 0 \end{pmatrix} \\
 U_{10}(x) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & [x]_1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad U_{11}(x) = \begin{pmatrix} 0 & 0 & [x]_1 \\ 0 & 0 & 0 \\ 0 & [x]_1 & 0 \end{pmatrix} \quad U_{12}(x) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & [x]_1 & 0 \end{pmatrix} \\
 U_{13}(x) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & [x]_1 \end{pmatrix} \\
 U_{14}(x) = \begin{pmatrix} [x]_2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad U_{15}(x) = \begin{pmatrix} 0 & [x]_2 & 0 \\ [x]_2 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad U_{16}(x) = \begin{pmatrix} 0 & 0 & [x]_2 \\ 0 & 0 & 0 \\ [x]_2 & 0 & 0 \end{pmatrix} \\
 U_{17}(x) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & [x]_2 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad U_{18}(x) = \begin{pmatrix} 0 & 0 & [x]_2 \\ 0 & 0 & 0 \\ 0 & [x]_2 & 0 \end{pmatrix} \quad U_{19}(x) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & [x]_2 \end{pmatrix}
 \end{array}$$

Die Koeffizientenvektoren für die Simulationsstudie in Kapitel 5.5 werden wie folgt festgelegt. **2 Kovariablen, 2 Zielgrößen, Kovarianztyp 1, Identitätslink:**

$$\beta = (5, 0.03, -0.77, 5, -0.29, 0.77)$$

$$\alpha = (-1.66, -1.1, -0.9, -0.91, 0.46, -0.28, 0.61, 0.27, -1.55)$$

2 Kovariablen, 2 Zielgrößen, Kovarianztyp 2, Identitätslink:

$$\beta = (5, 0.03, -0.77, 5, -0.29, 0.77)$$

$$\alpha = (-1.66, -1.1, -0.9, -0.91, 0.46, -0.28)$$

$$\Psi = \begin{pmatrix} 0.5 & 0 \\ 0 & 0.5 \end{pmatrix}$$

2 Kovariablen, 2 Zielgrößen, Kovarianztyp 3, Identitätslink:

$$\beta = (5, 0.03, -0.77, 5, -0.29, 0.77)$$

$$\alpha_{11} = (-1.66, -1.1, -0.9)$$

$$\alpha_{21} = (-0.91, 0.46, -0.28)$$

$$\alpha_{22} = (0.61, 0.27, -1.55)$$

4 Kovariablen, 2 Zielgrößen, Kovarianztyp 1, Logarithmuslink:

$$\beta = (7.00, 0.01, -0.19, -0.07, 0.19, 7.00, -0.41, -0.27, -0.23, -0.23)$$

$$\alpha = (0.35, -0.21, 0.45, 0.20, -1.16, 0.29, -0.43, -0.21, -1.34, -0.71, \\ -0.30, 1.01, 1.09, 0.35, 0.83)$$

4 Kovariablen, 2 Zielgrößen, Kovarianztyp 2, Logarithmuslink:

$$\beta = (7.00, 0.01, -0.19, -0.07, 0.19, 7.00, -0.41, -0.27, -0.23, -0.23)$$

$$\alpha = (0.35, -0.21, 0.45, 0.20, -1.16, 0.29, -0.43, -0.21, -1.34, -0.71)$$

$$\Psi = \begin{pmatrix} 0.5 & 0 \\ 0 & 0.5 \end{pmatrix}$$

4 Kovariablen, 2 Zielgrößen, Kovarianztyp 3, Logarithmuslink:

$$\beta = (7.00, 0.01, -0.19, -0.07, 0.19, 7.00, -0.41, -0.27, -0.23, -0.23)$$

$$\alpha_{11} = (0.46, -0.28, 0.61, 0.27, -1.55)$$

$$\alpha_{21} = (0.38, -0.57, -0.28, -1.79, -0.94)$$

$$\alpha_{22} = (-0.40, 1.34, 1.46, 0.46, 1.10)$$

4 Kovariablen, 3 Zielgrößen, Kovarianztyp 1, Identitätslink:

$$\beta = (5, 0.03, -0.77, -0.29, 0.77, 5, -1.66, -1.1, -0.9, -0.91, 5, 0.46, \\ -0.28, 0.61, 0.27)$$

$$\alpha = (-1.55, 0.38, -0.57, -0.28, -1.79, -0.94, -0.40, 1.34, 1.46, 0.46, \\ 1.10, -0.58, -0.38, 0.83, 1.35, -1.04, 1.08, -0.58, 0.14, -1.63, -1.32, \\ 1.60, -0.31, 0.99, 1.29, 1.82, 0.74, 0.00, -0.90, -1.08)$$

4 Kovariablen, 3 Zielgrößen, Kovarianztyp 2, Identitätslink:

$$\beta = (5, 0.03, -0.77, -0.29, 0.77, 5, -1.66, -1.1, -0.9, -0.91, 5, 0.46, \\ -0.28, 0.61, 0.27)$$

$$\alpha = (-1.55, 0.38, -0.57, -0.28, -1.79, -0.94, -0.40, 1.34, 1.46, 0.46, \\ 1.10, -0.58, -0.38, 0.83, 1.35)$$

$$\Psi = \begin{pmatrix} 0.5 & 0 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & 0.5 \end{pmatrix}$$

4 Kovariablen, 3 Zielgrößen, Kovarianztyp 3, Identitätslink:

$$\beta = (5, 0.03, -0.77, -0.29, 0.77, 5, -1.66, -1.1, -0.9, -0.91, 5, 0.46, \\ -0.28, 0.61, 0.27)$$

$$\alpha_{11} = (-1.55 \quad 0.38 \quad -0.57 \quad -0.28 \quad -1.79)$$

$$\alpha_{21} = (-0.94 \quad -0.4 \quad 1.34 \quad 1.46 \quad 0.46)$$

$$\alpha_{22} = (1.1 \quad -0.58 \quad -0.38 \quad 0.83 \quad 1.35)$$

$$\alpha_{31} = (-1.04 \quad 1.08 \quad -0.58 \quad 0.14 \quad -1.63)$$

$$\alpha_{32} = (-1.32 \quad 1.6 \quad -0.31 \quad 0.99 \quad 1.29)$$

$$\alpha_{33} = (1.82 \quad 0.74 \quad 0 \quad -0.9 \quad -1.08)$$

A.4. Datensätze

	F1	F2	F3	F4	F5	F6	F7	Y1	Y2
1	-1	-1	-1	-1	-1	-1	-1	711.90	0.59
2	-1	-1	-1	-1	-1	-1	-1	713.40	0.59
3	-1	-1	-1	-1	-1	-1	-1	712.30	0.47
4	-1	-1	-1	-1	-1	-1	-1	712.40	0.71
5	-1	-1	-1	-1	-1	-1	-1	711.90	0.63
6	-1	-1	-1	1	1	1	1	725.00	0.70
7	-1	-1	-1	1	1	1	1	720.10	0.91
8	-1	-1	-1	1	1	1	1	711.80	1.13
9	-1	-1	-1	1	1	1	1	723.90	0.79
10	-1	-1	-1	1	1	1	1	720.90	0.78
11	-1	1	1	-1	-1	1	1	711.60	0.56
12	-1	1	1	-1	-1	1	1	711.70	0.44
13	-1	1	1	-1	-1	1	1	711.30	0.46
14	-1	1	1	-1	-1	1	1	712.10	0.53
15	-1	1	1	-1	-1	1	1	711.70	0.46
16	-1	1	1	1	1	-1	-1	733.70	1.50
17	-1	1	1	1	1	-1	-1	724.10	1.55
18	-1	1	1	1	1	-1	-1	732.00	1.38
19	-1	1	1	1	1	-1	-1	732.70	1.45
20	-1	1	1	1	1	-1	-1	733.30	1.45
21	1	-1	1	-1	1	-1	1	725.40	1.25
22	1	-1	1	-1	1	-1	1	721.60	1.36
23	1	-1	1	-1	1	-1	1	722.60	1.51
24	1	-1	1	-1	1	-1	1	723.10	1.22
25	1	-1	1	-1	1	-1	1	721.10	1.25
26	1	-1	1	1	-1	1	-1	728.70	1.17
27	1	-1	1	1	-1	1	-1	721.10	0.97

28	1	-1	1	1	-1	1	-1	722.90	0.98
29	1	-1	1	1	-1	1	-1	723.00	0.97
30	1	-1	1	1	-1	1	-1	719.70	0.73
31	1	1	-1	-1	1	1	-1	726.60	1.52
32	1	1	-1	-1	1	1	-1	731.40	1.58
33	1	1	-1	-1	1	1	-1	731.40	1.61
34	1	1	-1	-1	1	1	-1	729.60	1.40
35	1	1	-1	-1	1	1	-1	731.30	1.57
36	1	1	-1	1	-1	-1	1	714.30	0.57
37	1	1	-1	1	-1	-1	1	714.40	0.51
38	1	1	-1	1	-1	-1	1	713.60	0.44
39	1	1	-1	1	-1	-1	1	716.30	0.44
40	1	1	-1	1	-1	-1	1	714.60	0.56

Tabelle A.3: Datensatz für die Anwendung in Kapitel 8.1

	L	K	D	FDV	Ve	Te	Wi	In	Po	Th
1	1	-1	1	-1	697.43	1574.70	9.45	17.29	4.70	353
2	1	1	1	1	763.42	1648.39	10.22	35.26	2.90	484
3	-1	-1	1	-1	671.12	1578.79	9.38	20.06	4.50	373
4	-1	-1	-1	1	682.18	1629.48	11.22	37.88	5.00	592
5	0	0	0	0	736.89	1645.51	9.45	31.13	3.20	415
6	0	0	0	0	731.55	1636.76	9.90	30.48	3.50	431
7	-1	1	1	-1	751.52	1645.71	10.31	23.17	2.80	294
8	-1	1	-1	1	757.36	1680.84	12.27	40.62	2.80	461
9	1	1	-1	1	773.23	1666.79	10.56	37.51	2.50	455
10	1	-1	-1	-1	703.23	1592.96	8.93	19.52	4.50	335
11	0	0	0	0	717.82	1621.22	10.03	27.09	4.10	422
12	-1	1	-1	-1	755.71	1662.92	9.18	24.09	3.00	284
13	1	1	-1	-1	770.80	1644.86	8.64	21.18	2.90	277

14	-1	1	1	1	724.02	1630.01	10.38	16.88	3.80	443
15	1	-1	1	1	669.14	1572.19	10.29	26.62	6.40	591
16	-1	-1	1	1	640.12	1571.78	10.38	28.31	6.80	600
17	-1	-1	-1	-1	676.38	1606.75	8.62	22.13	5.70	361
18	1	1	1	-1	753.64	1618.67	9.38	19.26	3.00	286
19	0	0	0	0	715.81	1621.80	9.29	27.55	3.70	429
20	1	-1	-1	1	691.73	1598.52	8.76	31.33	1.50	589
21	0	0	0	0	714.51	1618.62	9.36	26.63	2.50	437
22	0	0	-2	0	738.25	1656.00	8.65	31.91	2.00	406
23	-2	0	0	0	698.38	1621.21	9.23	28.33	2.80	452
24	2	0	0	0	738.57	1602.06	8.86	23.09	2.20	420
25	0	0	0	0	718.70	1608.03	8.97	24.80	2.20	410
26	0	0	0	-2	731.77	1609.80	8.67	13.76	5.10	206
27	0	0	2	0	687.26	1567.07	10.18	21.15	2.50	449
28	0	2	0	0	779.56	1655.71	8.86	28.92	1.30	355
29	0	0	0	2	707.12	1613.14	10.08	36.68	2.10	642
30	0	-2	0	0	629.32	1529.56	8.64	20.48	5.70	512

Tabelle A.4: Zentral zusammengesetzter Versuchsplan mit Partikel- und Schichtauswertungen

	L	K	D	FDV	Ve	Te	Wi	In
1	1	1	-1	-1	830.50	1786.74	7.33	22.33
2	1	-1	-1	1	725.59	1713.93	7.14	31.92
3	-1	-1	-1	-1	708.33	1722.78	6.60	21.61
4	-1	-1	1	1	660.93	1651.05	8.10	28.66
5	-1	1	1	-1	776.65	1743.66	8.38	21.72
6	1	1	1	1	787.77	1726.98	8.95	31.70
7	1	-1	1	-1	707.96	1632.23	8.02	15.47

8	-1	1	-1	1	795.94	1818.23	7.83	46.54
---	----	---	----	---	--------	---------	------	-------

Tabelle A.5: Teilfaktorieller initialer Versuchsplan für die HVOF-Anwendung in Kapitel 8.2 mit gemessenen Partikeleigenschaften

B. Notation

Die folgende Liste enthält die Erklärung der häufig verwendeten Symbole und Bezeichnungen in dieser Arbeit.

Symbol	Bezeichnung
--------	-------------

Mathematische Symbole

\mathbb{R} :	Menge der reellen Zahlen
\mathbb{N} :	Menge der natürlichen Zahlen

Bezeichnungen in Kapitel 3

$C \in \mathbb{R}^{p \times p}$	Kostenmatrix
$\tau \in \mathbb{R}^p$:	Vektor der Zielwerte
$loss(Y)$	Verlustfunktion
$R(x)$	Risikofunktion
$\mu(x)$	Bedingter Erwartungswertvektor $E(Y x)$
$\Sigma(x)$	Bedingte Kovarianzmatrix $Cov(Y x)$

Bezeichnungen in Kapitel 4

β :	Koeffizientenvektor im Erwartungswertmodell
η :	Linearer Prädiktor
$l(\cdot)$	Log-Likelihoodfunktion

Bezeichnungen in Kapitel 5

$K_h(\cdot)$:	Kernfunktion mit Bandbreite $h > 0$
$Z_1(x) \in \mathbb{R}^{q \times p}$:	Designmatrix im Erwartungswertmodell $\mu(x) = g^{-1}(Z_1(x)\beta)$
$A(x) \in \mathbb{R}^{p \times p}$:	Matrix-Logarithmus von $\Sigma(x)$
$U_j(x) \in \mathbb{R}^{p \times p}$:	Hilfsmatrizen zur Modellierung von $A(x)$
$\Psi \in \mathbb{R}^{p \times p}$:	Positiv definite Matrix als Teil des Kovarianzregressionsmodells

$Z_2(x) \in \mathbb{R}^{w \times p}$:	Designmatrix im Kovarianzregressionsmodell $\Sigma(x) = \Psi + (Z_2(x)\alpha) \cdot (Z_2(x)\alpha)^T$
--	--

Bezeichnungen in Kapitel 6

$R_{vec}(x)$:	Vektor der Einzelrisiken
$A_Y \in \mathbb{R}^{p \times p}$:	Standardisierungsmatrix
$W \in \mathbb{R}^{p \times p}$:	Gewichtsmatrix

Des Weiteren werden im folgenden die dieser Arbeit zugrundeliegenden Notationen erläutert.

Mit $Y \in \mathbb{R}^p$ wird immer der Zufallsvektor der Zielvariablen und mit $x \in \mathbb{R}^k$ wird immer der Kovariablenvektor bezeichnet. Der Fall $p = 1$ wird in Kapitel 4 behandelt, während die darauf folgenden Kapitel sich auf den allgemeinen Fall $p \in \mathbb{N}$ beziehen.

Der Eintrag in Zeile i und Spalte j einer Matrix $A \in \mathbb{R}^{k_1, k_2}$ wird durch $[A]_{i,j}$ dargestellt. Der i -te Eintrag eines Vektors $a \in \mathbb{R}^{k_1}$ wird mit $[a]_i$ bezeichnet. Die i -te Zeile der Matrix A wird mit $[A]_{i, \cdot}$ bezeichnet und die j -te Spalte wird mit $[A]_{\cdot, j}$ dargestellt.

Der Ausdruck A^T oder y^T meint die Transponierte einer Matrix A oder eines Vektors y . Ein Vektor $y \in \mathbb{R}^p$ ist immer ein Spaltenvektor und entsprechend ist y^T ein Zeilenvektor. Ferner wird mit $\det(A)$ die Determinante und mit $tr(A)$ die Spur einer Matrix A bezeichnet.

Die Schätzung einer Zufallsvariable Y wird immer mit \hat{Y} bezeichnet.