

**Fakultät  
Statistik**

# **Bachelorarbeit**

über das Thema

**Vergleich der Prognosegüte verschiedener Regressionsmodelle für  
longitudinale Daten**

**Autor:** Gerrit Toenges  
gerrit.toenges@tu-dortmund.de

**Prüferin:** Prof. Dr. Christine Müller

**Abgabedatum:** 19.12.2013

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
<b>2</b>	<b>Datenbeschreibung und Zielsetzung</b>	<b>4</b>
2.1	Einführung in die Struktur longitudinaler Daten . . . . .	4
2.2	Datenbeschreibung . . . . .	5
2.3	Zielsetzung . . . . .	6
<b>3</b>	<b>Statistische Methoden</b>	<b>8</b>
3.1	Lineare Interpolation . . . . .	8
3.2	Verteilungstheorie . . . . .	9
3.2.1	Die multivariate Normalverteilung und ihre Folgeverteilungen . . . . .	9
3.2.2	Verteilung quadratischer Formen . . . . .	10
3.3	Das multiple lineare Regressionsmodell . . . . .	11
3.3.1	Das klassische lineare Regressionsmodell . . . . .	12
3.3.1.1	Parameterschätzung . . . . .	13
3.3.1.2	Punkt- und Intervallprognosen für zukünftige Beobachtungen . . . . .	18
3.3.2	Das verallgemeinerte lineare Regressionsmodell . . . . .	21
3.4	Das Regressionsmodell mit zufälligen Effekten (RCR-Modell) . . . . .	22
3.4.1	Parameterschätzung im RCR-Modell . . . . .	25
3.4.1.1	Schätzung der festen Effekte . . . . .	26
3.4.1.2	Schätzung der Varianzparameter . . . . .	30
3.4.2	Prädiktion der zufälligen Effekte . . . . .	34
3.4.3	Prognosen und Prognoseintervalle für zukünftige Beobachtungen im RCR-Modell . . . . .	39
3.4.3.1	Situation A: Prognosen für einen neuen, unbekanntem Verlauf . . . . .	39
3.4.3.2	Situation B: Prognosen für einen vollständig beobachteten Verlauf . . . . .	43
3.4.3.3	Situation C: Prognosen für einen neuen Verlauf, von dem nur die ersten Messungen bekannt sind . . . . .	43
<b>4</b>	<b>Statistische Auswertung</b>	<b>44</b>
4.1	Datenaufbereitung durch lineare Interpolation . . . . .	44
4.2	Deskriptive Auswertung der Dialyseverläufe . . . . .	45
4.3	Empirischer Vergleich verschiedener Prognoseverfahren . . . . .	48
4.3.1	Verfahren 1: Prognose des Endpunktes eines neuen Verlaufs ausschließlich auf Basis vorheriger Verläufe . . . . .	50

4.3.2	Verfahren 2 und 3: Prognose des Endpunktes eines neuen Ver- laufs unter Hinzunahme von dessen Anfangsbeobachtungen . . .	53
4.3.3	Verfahren 4: Prognosen basierend auf einem Modell mit der Sig- nalstärke als Regressor . . . . .	57
4.3.4	Vergleich der vier Prognoseverfahren . . . . .	62
<b>5</b>	<b>Zusammenfassung und Ausblick</b>	<b>65</b>
	<b>Anhang</b>	<b>68</b>
	<b>A Verlaufprofile aller Patienten</b>	<b>68</b>
	<b>B R-Code</b>	<b>72</b>
	<b>Literaturverzeichnis</b>	<b>94</b>
	<b>Eidesstattliche Versicherung</b>	<b>97</b>

# 1 Einleitung

Die Nieren sind ein paarig angelegtes Organ, welches sich auf der rechten und linken Seite der unteren Wirbelsäule befindet. In den Nieren werden toxische Substanzen (Stoffwechselendprodukte, Medikamente etc.) aus dem Blut abfiltriert und in wenig Wasser aufkonzentriert. Dieses Konzentrat wird letztlich über das Harnwegesystem ausgeschieden. Die Niere hat somit eine zentrale Funktion für die Entgiftung und den Wasserhaushalt des Körpers. Bei einer Niereninsuffizienz werden die Toxine nicht mehr in ausreichendem Maße entfernt und reichern sich im Körper an, was sich negativ auf alle Organsysteme auswirkt. Man unterscheidet dabei chronisches und akutes Nierenversagen. Unter chronischem Nierenversagen fasst man den langsam fortschreitenden Funktionsverlust der Niere zusammen, der meist die Folge einer chronischen Ersterkrankung wie z.B. Diabetes mellitus ist. Beim akuten Nierenversagen tritt ein plötzlicher Funktionsverlust des Organs infolge einer akuten, schweren Ersterkrankung wie z.B. einer Blutvergiftung ein (Schreiber et al., 2009, S. 1525 ff.).

Im Jahr 2005 waren in Deutschland rund 55.000 Menschen von einer Niereninsuffizienz betroffen. Bei diesen Patienten kann die Nierenfunktion – zumindest teilweise – durch ein technisches Blutreinigungsverfahren, die Dialyse, ersetzt werden. Dabei wird das Blut des Patienten durch einen Schlauch in eine Dialysemaschine geleitet, in der eine Entgiftung des Blutes stattfindet. Das gereinigte Blut wird dem Patienten anschließend wieder zugeführt. Für die Dialyse wird also ein extrakorporaler Blutkreislauf eingerichtet. Chronisch dialysepflichtige Patienten müssen drei mal pro Woche zur Blutreinigung erscheinen. Die Konzentration an Toxinen, die sich zwischen zwei Dialysebehandlungen im Körper anreichert, wird maßgeblich durch die Art der Ernährung in diesem Zeitraum mitbestimmt (Schreiber et al., 2009, S. 1519 ff.).

Das Herzstück des extrakorporalen Blutkreislaufs ist der Dialysator. Die in diesem Gerät stattfindende Blutwäsche basiert auf dem physikalischen Prinzip der Diffusion. Die Diffusion beschreibt den Vorgang des Konzentrationsausgleichs zwischen zwei unterschiedlich konzentrierten Lösungen. Es findet so lange eine Netto-Teilchenwanderung vom Ort hoher zum Ort niedriger Konzentration statt, bis beide Lösungen die gleiche Konzentration erreicht haben. Dieser Vorgang kann selektiv gestaltet werden, indem man die beiden Lösungen durch eine semipermeable Membran trennt, welche nur den Durchtritt kleiner Teilchen erlaubt. Im Fall der Dialyse handelt es sich bei den beiden Lösungen um das Blut und eine Dialysierflüssigkeit, die eine physiologische Gesamtkonzentration gelöster Salze enthält (Schreiber et al., 2009, S. 1534 ff.).

Der Dialysator besteht aus einem Kunststoffgehäuse, durch welches tausende, nur wenige Mikrometer breite Hohlfasern verlaufen. Die Wand dieser kleinen Röhren besteht aus einer semipermeablen Membran. Während das Patientenblut durch die Hohl-

fasern fließt, strömt an der Außenseite der Fasern die Dialysierflüssigkeit in entgegengesetzter Richtung entlang. Da in der Dialysierflüssigkeit keine Toxine vorhanden sind, treten die Stoffwechselendprodukte aus dem Blut in die Dialysierflüssigkeit über. Die verbrauchte Dialysierflüssigkeit – das Dialysat – weist nach Passage des Dialysators dieselbe Toxinkonzentration auf wie das gereinigte Blut und wird entsorgt. Die pro Zeiteinheit entfernte Toxinmenge ist abhängig von der Geschwindigkeit bzw. der Flussrate, mit der das Blut und die Dialysierflüssigkeit durch den Dialysator strömen. Je höher die Flussrate ist, desto schneller werden die Toxine aus dem Blut entfernt. Dennoch sind kurze Dialysezeiten mit hoher Flussrate nicht sinnvoll, da die im Gewebe lagernden Gifte nur langsam an das Blut abgegeben werden. Um eine Entgiftung aller Körperkompartimente zu erzielen, muss eine minimale Behandlungsdauer von ca. vier Stunden eingehalten werden (Geberth, 2011, S. 30 ff.).

Der Entgiftungsprozess wird durch Beobachtung der Reduktion von Harnstoff im Blut überwacht. Das Toxin Harnstoff ist ein Endprodukt des Eiweißstoffwechsels und wird stellvertretend für alle kleinemolekularen Toxine zur Labordiagnostik herangezogen. Ein wichtiges Maß zur Beurteilung der Dialyseeffektivität ist die Harnstoffreduktionsrate

$$URR := 1 - \frac{C_T}{C_0}.$$

Hierbei ist  $C_0$  die Harnstoffkonzentration im Blut vor und  $C_T$  die Harnstoffkonzentration im Blut nach der Dialyse. Die Harnstoffreduktionsrate gibt also an, um wieviel Prozent die ursprüngliche Harnstoffmenge während einer Dialysebehandlung gesenkt wurde. Der empfohlene Zielwert liegt bei einer  $URR$  von ca. 65% (Geberth, 2011, S. 117 ff.).

Bis vor einigen Jahren wurden für die Bestimmung dieser Größe jeweils vor und nach der Behandlung Blutproben entnommen und im Labor auf ihre Harnstoffkonzentration hin untersucht. Das Problem dabei war, dass die Dialyseeffektivität mit diesem Verfahren immer erst nach Beendigung der Behandlung beurteilt werden konnte. Seit einiger Zeit gibt es jedoch Dialysegeräte, die die Reduktion der Harnstoffkonzentration in Echtzeit mitverfolgen können. In diesen Maschinen wird die Harnstoffkonzentration des aus dem Dialysator austretenden Dialysats gemessen, da diese der Konzentration im gereinigten Blut entspricht (Uhlen et al., 2005).

Technisch wird diese Messung durch ein Spektrophotometer realisiert: Dabei wird von einer UV-Lampe ultraviolettes Licht durch das Dialysat gestrahlt. Auf der gegenüberliegenden Seite wird die Intensität des ankommenden Lichts durch einen Detektor gemessen. Da die niedermolekularen Toxine UV-Licht absorbieren, kann aus der Intensität der transmittierten Reststrahlung ein Rückschluss auf die Toxinkonzentration im Dialysat gezogen werden. Anhand der Intensität kann die Absorbanz des Dialysats

bestimmt werden. Die Absorbanz  $A$  ist gegeben durch

$$A := \log \left( \frac{I_r}{I_s} \right) .$$

Hierbei ist  $I_r = 55.000$  die Intensität des transmittierten Lichts durch reine Dialysierflüssigkeit und  $I_s$  die Intensität des transmittierten Lichts durch verbrauchtes Dialysat. Weil die Absorbanz eine hohe Korrelation zur Harnstoffkonzentration aufweist, repräsentiert die Absorbanz die Harnstoffkonzentration im Dialysat und damit im gereinigten Blut (Uhlin et al., 2006).

Durch den im Laufe der Dialyse fortschreitenden Harnstoffentzug sinkt die Harnstoffkonzentration im Blut und im Dialysat. Die gemessenen Intensitätswerte des Lichtdetektors steigen während der Behandlung stetig an, da immer weniger Harnstoffmoleküle das Licht absorbieren. Mit diesem Verfahren kann zu jedem Zeitpunkt der Behandlung die bislang erreichte Harnstoffreduktionsrate berechnet werden. Diese Methode ermöglicht es also, die Dialyse so lange fortzusetzen, bis die nötige Harnstoffreduktionsrate erreicht ist (Uhlin et al., 2005).

Sowohl für das Pflegepersonal als auch für den Patienten wäre es wünschenswert, zu Beginn der aktuellen Behandlung eine Prognose zu erhalten, wie lange die bevorstehende Sitzung dauern wird. Anhand einer solchen Prognose können sich die Patienten besser auf die Behandlung vorbereiten und das Pflegepersonal kann die Auslastung der Geräte besser planen. Im Rahmen dieser Bachelorarbeit werden verschiedene Prognoseverfahren verglichen, mit denen die Harnstoffkonzentration zu zukünftigen Behandlungszeitpunkten für jeden Patienten individuell vorhergesagt werden kann. Als Zielgröße wird die vom Detektor gemessene Lichtintensität modelliert. Diese kann dann mithilfe der vorgestellten Formeln in die Absorbanz und die Harnstoffreduktionsrate umgerechnet werden.

Zunächst wird in Kapitel 2 die allgemeine Struktur longitudinaler Daten erläutert. Es folgt eine Beschreibung der vorliegenden Dialysedaten und eine genaue Formulierung der Zielsetzung dieser Bachelorarbeit. Anschließend werden in Kapitel 3 die statistischen Regressionsmodelle vorgestellt, welche die Grundlage der verschiedenen Prognoseverfahren sind. In Kapitel 4 werden die Prognoseverfahren auf die Dialyseverläufe angewendet und hinsichtlich ihrer Vorhersagegenauigkeit miteinander verglichen. Zuletzt folgt in Kapitel 5 eine Zusammenfassung der wesentlichen Ergebnisse dieser Arbeit.

## 2 Datenbeschreibung und Zielsetzung

### 2.1 Einführung in die Struktur longitudinaler Daten

Longitudinale Daten bestehen aus einer Zufallsstichprobe von Subjekten aus einer Population. Für jedes Subjekt wird eine bestimmte Variable über die Zeit hinweg beobachtet. Die Begriffe Subjekt und Population stammen aus der wirtschaftswissenschaftlichen Anwendung, wo viele Individuen aus der Bevölkerung über die Zeit hinweg untersucht werden. Im Rahmen dieser Bachelorarbeit werden verschiedene Dialyseverläufe einer Person modelliert. Daher werden im Folgenden ausschließlich die Begriffe „Verlauf“ oder „Beobachtungsreihe“ verwendet. Ein Verlauf ist die Realisation einer experimentellen Einheit, d.h. einer einzelnen Dialysebehandlung.

Modelle für longitudinale Daten unterscheiden sich von Regressions- und Zeitreihenmodellen durch ihre Doppelindizes. Diese ermöglichen es, zwischen den verschiedenen Verläufen und den Zeitpunkten zu unterscheiden. Dabei ist  $y_{ij}$  die Realisation der Zufallsvariablen  $Y_{ij}$ , welche die Verteilung der abhängigen Variablen des  $i$ -ten Verlaufs zum  $j$ -ten Zeitpunkt  $t_{ij}$  beschreibt. Ein longitudinaler Datensatz enthält die Zeitreihen von  $I$  experimentellen Einheiten. Die Gesamtzahl der Messzeitpunkte von Verlauf  $i$  wird mit  $N_i$  gekennzeichnet. Falls die Anzahl der Beobachtungen pro Beobachtungsreihe von  $i$  abhängt – d.h. wenn unterschiedlich viele Messungen pro Verlauf durchgeführt werden – werden die Daten als unbalanciert bezeichnet. Im Falle  $N_i = N \forall i \in \{1, \dots, I\}$  liegen balancierte Daten vor. Ein Spezialfall balancierter Longitudinaldaten ist gegeben, wenn die  $N$  Beobachtungszeitpunkte aller Verläufe gleich sind, d.h. wenn  $t_{ij} = t_j \forall i \in \{1, \dots, I\}$  erfüllt ist (Frees, 2006, S. 4).

Unbalancierte Daten resultieren in der Regel aus Beobachtungsstudien, in denen die Messzeitpunkte nicht genau kontrolliert werden können. Bei experimentellen Studien hingegen wird das Studiendesign meist so angelegt, dass balancierte Longitudinaldaten mit gleichen, äquidistanten Zeitpunkten vorliegen. Der Verlauf einer Beobachtungseinheit wird auch als Profil bezeichnet. Bei balancierten Daten mit gleichen Zeitpunkten liegen also  $I$  Profile der Form  $\{(t_1, Y_{i1}), \dots, (t_N, Y_{iN})\}$ ,  $i \in \{1, \dots, I\}$ , vor (Weiss, 2005, S. 16 ff.).

Das Ziel der longitudinalen Datenanalyse ist, die Verläufe aller experimentellen Einheiten gemeinsam zu modellieren und Gesetzmäßigkeiten für die zeitliche Dynamik der Gesamtpopulation, d.h. eines „Durchschnittsverlaufs“, zu erforschen. Auf dieser Basis können dann Vorhersagen für die Verläufe neuer Einheiten gemacht werden (Weiss, 2005, S. 2 ff.). Longitudinale Daten haben eine breite Anwendung in vielen Bereichen der Natur-, Wirtschafts- und Sozialwissenschaften. In den Wirtschafts- und Sozialwissenschaften sind longitudinale Daten unter dem Begriff „Panel“-Daten verbreit-

tet, während sie in der Biometrie eher als „Wachstumskurven“-Daten bekannt sind. Trotz der diversen synonymen Bezeichnungen ist die statistische Methodik in allen Anwendungsgebieten identisch (Frees, 2006, S. 4 ff.).

## 2.2 Datenbeschreibung

Für diese Bachelorarbeit werden Daten aus Dialysebehandlungen von insgesamt 21 dialysepflichtigen Patienten ausgewertet. Die Daten stammen aus klinischen Studien der B. Braun Melsungen AG und wurden in den Jahren 2007 bis 2009 in den Dialysekliniken in Melsungen und Bad Wildungen erhoben. Die B. Braun Melsungen AG ist ein Medizintechnik-Unternehmen, welches maßgeblich an der Entwicklung eines Systems zur UV-basierten Konzentrationsbestimmung harnpflichtiger Substanzen im Dialysat beteiligt war. Die Patienten sind im Datensatz – aus unbekanntem, möglicherweise Klinik-internen Gründen – nicht fortlaufend nummeriert. Für die Auswertung der Daten werden die Patienten in eine fortlaufende Reihenfolge unnummeriert (siehe Tabelle 1).

Für jede Dialysebehandlung liegen die Messzeitpunkte (in Minuten nach Beginn der Dialysesitzung) sowie die zu diesen Zeiten gemessenen Signalstärken des UV-Detektors vor. Jede Behandlung wird also durch den zeitlichen Verlauf der quantitativ stetig skalierten Intensitätswerte des Detektors repräsentiert. Die Anzahl der Verläufe pro Patient variiert im Datensatz: Es liegen Daten zu maximal 32 und minimal 8 Dialysebehandlungen pro Patient vor. Die klinischen Daten bestehen also aus 21 Patientenspezifischen longitudinalen Datensätzen.

Die Dauer einer Dialysebehandlung ist von Patient zu Patient und sogar von Verlauf zu Verlauf innerhalb eines Patienten unterschiedlich, liegt jedoch meistens über 230 Minuten. Da sich die Anzahl der Messzeitpunkte pro Behandlung selbst innerhalb eines Patienten unterscheidet, sind die 21 longitudinalen Datensätze jeweils unbalanciert. Zudem weisen die Verläufe jedes Patienten unterschiedliche Messzeitpunkte auf: Die erste Messung erfolgte in den meisten Fällen zwischen den ersten 3 bis 5 Minuten nach Dialysebeginn, da erst dann ein Konzentrationsausgleich zwischen der Dialysierflüssigkeit und dem Blut vorliegt. Die darauf folgenden Messungen wurden in nicht äquidistanten Zeitintervallen von ca. 3 bis 4 Minuten durchgeführt, sodass meist über 70 Beobachtungen pro Verlauf vorliegen. Eine detaillierte Auflistung der Patienten sowie der Anzahl und Charakteristika ihrer Verläufe ist in Tabelle 1 gegeben.



Hinweis:

In diesem Auszug der Bachelorarbeit ist Tabelle 1 aus Datenschutzgründen nicht enthalten.

## 2.3 Zielsetzung

Das Ziel dieser Bachelorarbeit ist, ein Verfahren zur Prognose der Signalstärke des Detektors am Ende einer Dialysebehandlung zu entwickeln. Die Prognose sollte entweder bereits vor der Behandlung oder im Anfangsstadium der Behandlung vorliegen. Dazu werden verschiedene Prognoseverfahren miteinander verglichen, die auf klassischen Regressionsmodellen sowie auf Regressionsmodellen mit zufälligen Effekten beruhen. Es werden sowohl Punktprognosen als auch Prognoseintervalle für die Signalintensität am Ende der Dialyse konstruiert.

Da die Dauer einer Dialysebehandlung immer unterschiedlich ist und die Prognosegüte der einzelnen Verfahren am Ende Patienten-übergreifend verglichen werden soll, wird der Detektorwert jeweils immer für den Zeitpunkt  $T = 235$  prognostiziert. Dies hat den Grund, dass die meisten Behandlungen der im Datensatz vertretenen Patienten ca. 4 Stunden dauern. Nur wenige Patienten weisen deutlich längere Zeiten auf. Durch

eine Prognose des Detektorwertes für den Endzeitpunkt dieser Standard-Dialysedauer kann zu Beginn der Behandlung eingeschätzt werden, ob für die aktuelle Sitzung eine Behandlungsdauer von 4 Stunden ausreicht.

Bei einigen Verfahren wird die Information vorheriger Verläufe eines Patienten für die Prognose von dessen aktuellem Verlauf verwendet. Aus diesem Grund werden alle Prognoseverfahren jeweils nur zur Prognose des (im Datensatz) letzten Verlaufs jedes Patienten angewendet. Die Evaluation der Prognosegüte aller Verfahren erfolgt also anhand des Vergleichs zwischen der Prognose und der Beobachtung des letzten Verlaufs jedes Patienten zum Zeitpunkt  $T = 235$ .

## 3 Statistische Methoden

In diesem Kapitel werden die Regressionsmodelle, welche die Basis der verschiedenen Prognoseverfahren sind, vorgestellt. Dabei handelt es sich um das multiple lineare Regressionsmodell (Abschnitt 3.3) und um das Regressionsmodell mit zufälligen Effekten (Abschnitt 3.4). Einige Grundlagen aus der Verteilungstheorie, die insbesondere zur Herleitung eines Prognoseintervalls im klassischen linearen Modell benötigt werden, werden kurz in Abschnitt 3.2 eingeführt. Zunächst wird jedoch in Abschnitt 3.1 ein nicht-statistisches Verfahren, die lineare Interpolation, behandelt. Diese Methode wird später benötigt, um die einzelnen Beobachtungsreihen so aufzubereiten, dass die statistischen Modelle angewendet werden können.

### 3.1 Lineare Interpolation

Die Methode der linearen Interpolation wird in dieser Arbeit verwendet, um das Problem der ungleichen Messzeitpunkte in den Dialysedaten zu umgehen. Für die Auswertung der Daten wird jede Beobachtungsreihe linear interpoliert, um anschließend alle Verläufe an denselben Messzeitpunkten auszuwerten. Durch diese Transformationen der Original-Zeitreihen erhält man balancierte Longitudinaldaten mit gleichen Zeitpunkten. Diese Datenstruktur lässt sich wesentlich einfacher modellieren. Im Folgenden wird das Verfahren allgemein beschrieben.

Ein Interpolationsproblem besteht darin, zu  $N$  Datenpunkten  $(x_1, y_1), \dots, (x_N, y_N) \in \mathbb{R}^2$ ,  $x_1 < \dots < x_N$ , eine Funktion  $f : \mathbb{R} \rightarrow \mathbb{R}$  zu finden, die die Bedingung  $f(x_i) = y_i \forall i \in \{1, \dots, N\}$  erfüllt. Da unendlich viele Funktionen existieren, die diese Bedingung erfüllen, wird die funktionale Form von  $f$  vorab durch ein parametrisches Modell festgelegt. Die Parameter können dann anhand der Daten eindeutig bestimmt werden (Werner, 1992, S. 135 ff.).

Bei der linearen Interpolation werden benachbarte Datenpunkte jeweils stückweise linear miteinander verbunden, sodass eine stetige Funktion  $f$  entsteht, die auf dem Intervall  $[x_1, x_N]$  definiert ist. Die Funktion  $f$  kann nun an einer beliebigen Stelle  $x_0 \in [x_1, x_N]$  ausgewertet werden: Seien dazu  $x_i$  und  $x_{i+1}$ ,  $i \in \{1, \dots, N-1\}$ , die Stützstellen der Datenpunkte, für die  $x_i \leq x_0 \leq x_{i+1}$  gelte. Dann ist der Funktionswert von  $f$  an der Stelle  $x_0$  gegeben durch

$$y_0 = f(x_0) = \frac{y_{i+1} - y_i}{x_{i+1} - x_i} (x_0 - x_i) + y_i.$$

Die Interpolation dient also genau wie die Regression dazu, einen funktionellen Zusammenhang zwischen zwei Variablen herzustellen und neue Werte zu schätzen. Im

Gegensatz zur Regression werden die Datenpunkte bei der Interpolation aber nicht als Ergebnis eines Zufallsprozesses angesehen: Alle Datenpunkte müssen auf der interpolierenden Funktion liegen (Werner, 1992, S. 135 ff.).

## 3.2 Verteilungstheorie

Die Theorie der Regressionsmodelle erfordert den Umgang mit Zufallsvektoren, d.h. mit multivariaten Datenstrukturen. Die bekannteste multivariate Verteilung, die den Beobachtungen in Regressionsmodellen unterstellt wird, ist die multivariate Normalverteilung. Von dieser Verteilung lassen sich zahlreiche weitere Verteilungen ableiten, die bei der Konstruktion von Prognoseintervallen eine wichtige Rolle spielen. Im Folgenden werden die Grundlagen der in den weiteren Abschnitten benötigten Verteilungstheorie vorgestellt.

### 3.2.1 Die multivariate Normalverteilung und ihre Folgeverteilungen

#### Definition 3.1: Multivariate Normalverteilung

Ein  $(N \times 1)$ -dimensionaler Zufallsvektor  $\mathbf{Y} := (Y_1, \dots, Y_N)'$  heißt nach Seber (2007, S. 435 f.) multivariat normalverteilt, wenn für jedes  $\mathbf{a} \in \mathbb{R}^N \setminus \{\mathbf{0}\}$  gilt, dass die Zufallsvariable  $Z := \mathbf{a}'\mathbf{Y}$  univariat normalverteilt ist. Eine multivariate Normalverteilung ist durch den Erwartungswertvektor  $\boldsymbol{\mu} \in \mathbb{R}^N$  und die positiv definite Kovarianzmatrix  $\mathbf{V} \in \mathbb{R}^{N \times N}$  eindeutig definiert. Die Dichte von  $\mathbf{Y} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{V})$  ist gegeben durch

$$f_{\mathbf{Y}}(y_1, \dots, y_N) = \left( \frac{1}{\sqrt{2\pi}} \right)^N \frac{1}{\sqrt{\det(\mathbf{V})}} \exp \left( -\frac{1}{2}(\mathbf{y} - \boldsymbol{\mu})' \mathbf{V}^{-1}(\mathbf{y} - \boldsymbol{\mu}) \right).$$

#### Satz 3.2: Unabhängigkeit zweier multivariat normalverteilter Zufallsvektoren

Seien  $\mathbf{Y}_1 := (Y_{11}, \dots, Y_{1k})'$  und  $\mathbf{Y}_2 := (Y_{21}, \dots, Y_{2f})'$  zwei multivariat normalverteilte Zufallsvektoren, für die auch  $\mathbf{Y} := (\mathbf{Y}_1', \mathbf{Y}_2')'$  multivariat normalverteilt ist. Dann gilt:  $\mathbf{Y}_1$  und  $\mathbf{Y}_2$  sind genau dann stochastisch unabhängig, wenn  $\text{Cov}(\mathbf{Y}_1, \mathbf{Y}_2) = \mathbf{0}$  ist. Hierbei bezeichnet  $\mathbf{0}$  eine  $(k \times f)$ -Matrix, die nur aus Nullen besteht (Casparly und Wichmann, 1994, S. 182).

#### Definition 3.3: $\chi^2$ -Verteilung

Sei  $\mathbf{Y} := (Y_1, \dots, Y_N)' \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{I}_N)$ , wobei  $\mathbf{I}_N$  die  $(N \times N)$ -Einheitsmatrix bezeichnet. Dann ist  $\mathbf{Y}'\mathbf{Y}$   $\chi^2$ -verteilt mit  $N$  Freiheitsgraden und Nichtzentralitätsparameter  $\delta = \boldsymbol{\mu}'\boldsymbol{\mu}$ . Es gilt die Notation

$$\mathbf{Y}'\mathbf{Y} \sim \chi_{N, \delta}^2.$$

Falls  $\delta = 0$  ist, liegt eine zentrale  $\chi^2$ -Verteilung mit  $N$  Freiheitsgraden vor (Casparly und Wichmann, 1994, S. 184 f.).

**Definition 3.4:  $t$ -Verteilung**

Seien  $Z \sim \mathcal{N}(0,1)$  und  $W \sim \chi_N^2$  stochastisch unabhängig. Dann heißt die Verteilung von

$$T := \frac{Z}{\sqrt{\frac{1}{N}W}}$$

$t$ -Verteilung mit  $N$  Freiheitsgraden. Es gilt die Notation  $T \sim t_N$  (Casparly und Wichmann, 1994, S. 186).

### 3.2.2 Verteilung quadratischer Formen

**Definition 3.5: Quadratische Form**

Sei  $x \in \mathbb{R}^N$  ein Vektor und  $A \in \mathbb{R}^{N \times N}$  eine symmetrische Matrix. Dann heißt die Abbildung  $q(x) := x'Ax$  quadratische Form (Casparly und Wichmann, 1994, S. 47).

**Definition 3.6: Idempotente Matrix**

Eine quadratische Matrix  $A \in \mathbb{R}^{N \times N}$  ist idempotent, falls  $AA = A$  gilt (Toutenburg, 1992, S. 282).

**Satz 3.7: Rang einer idempotenten Matrix**

Für eine idempotente Matrix  $A \in \mathbb{R}^{N \times N}$  gilt stets  $\text{rang}(A) = \text{spur}(A)$  (Toutenburg, 1992, S. 282).

**Satz 3.8:  $\chi^2$ -Verteilung einer quadratischen Form**

Sei  $Y := (Y_1, \dots, Y_N)' \sim \mathcal{N}(\mu, I_N)$  ein stochastischer Vektor und  $A \in \mathbb{R}^{N \times N}$  eine symmetrische Matrix. Dann gilt: Die quadratische Form  $Y'AY$  ist genau dann  $\chi_{r,\delta}^2$  mit Nichtzentralitätsparameter  $\delta = \mu' A \mu$ , wenn  $A$  idempotent und  $\text{rang}(A) = r$  ist (Casparly und Wichmann, 1994, S. 192).

### 3.3 Das multiple lineare Regressionsmodell

Lineare Modelle nehmen eine wichtige Stellung in den modernen statistischen Methoden ein. Das multiple lineare Regressionsmodell stellt einen linearen funktionalen Zusammenhang zwischen einer metrisch skalierten Zielgröße  $Y$  und  $(k - 1)$  Einflussgrößen  $X_1, \dots, X_{k-1}$  her. Die Zielgröße wird als abhängige Variable und die Einflussgrößen werden als unabhängige Variablen oder Regressoren bezeichnet (Toutenburg, 1992, S. 21).

Bei longitudinalen Daten dient meist nur die Zeit als Regressor. Auf diese Weise kann jede Beobachtungsreihe einzeln durch ein lineares Modell modelliert werden. Die Kenntnis früherer Verläufe fließt hier nicht in die Modellierung mit ein. Eine weitere Anwendungsmöglichkeit des linearen Modells bei longitudinalen Daten besteht darin, den Wert der abhängigen Variablen zu einem späten Beobachtungszeitpunkt in Abhängigkeit des Wertes zu einem frühen Beobachtungszeitpunkt zu modellieren. In diesem Fall wird also nicht die Zeit als Regressor genutzt und die Information aller Verläufe geht in die Modellierung mit ein.

Die Regressionsanalyse basiert auf  $N$  Beobachtungsvektoren, welche von der Form  $(y_i, x_{i1}, \dots, x_{i(k-1)})'$ ,  $i \in \{1, \dots, N\}$ , sind. Dabei ist  $y_i$  die Ausprägung der Zielgröße, die gemeinsam mit den Ausprägungen  $x_{i1}, x_{i2}, \dots, x_{i(k-1)}$  der unabhängigen Variablen realisiert wird. Im linearen Regressionsmodell lässt sich eine Beobachtung der abhängigen Variablen  $Y$  approximativ als Linearkombination der beobachteten Werte der unabhängigen Variablen beschreiben (Toutenburg, 1992, S. 18 ff.):

$$Y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_{k-1} x_{i(k-1)} + \epsilon_i = \mathbf{x}_i' \boldsymbol{\beta} + \epsilon_i, \quad i = 1, \dots, N.$$

Hierbei ist  $\boldsymbol{\beta} := (\beta_0, \dots, \beta_{k-1})'$  der  $(k \times 1)$ -Vektor der unbekanntenen Regressionskoeffizienten,  $\mathbf{x}_i := (1, x_{i1}, \dots, x_{i(k-1)})'$  der  $(k \times 1)$ -Vektor der Beobachtungen der unabhängigen Variablen und  $\epsilon_i$  eine stochastische Störgröße mit Erwartungswert 0.

Da es weitere, nicht im Modell befindliche Variablen gibt, die einen Einfluss auf die Zielgröße ausüben, kann der lineare Zusammenhang zwischen der abhängigen Variablen und den unabhängigen Variablen nur als approximativ gültig angesehen werden. Der Einfluss unbekannter Variablen wird durch die stochastische Störgröße  $\epsilon_i$  modelliert. Durch den stochastischen Charakter der Störgröße wird auch die Beobachtung  $Y_i$  zu einer Zufallsvariablen (Urban, 2006, S. 35 ff.).

Die  $N$  linearen, stochastischen Gleichungen des multiplen linearen Regressionsmo-

dells lassen sich kompakt in Matrixform formulieren als

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_N \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \dots & x_{1(k-1)} \\ 1 & x_{21} & \dots & x_{2(k-1)} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{N1} & \dots & x_{N(k-1)} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{k-1} \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_N \end{bmatrix} \iff \mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}.$$

Hierbei ist  $\mathbf{Y} := (Y_1, \dots, Y_N)'$  der  $(N \times 1)$ -Vektor der Beobachtungen der abhängigen Variablen,  $\mathbf{X}$  die  $(N \times k)$ -Designmatrix mit den Beobachtungen der unabhängigen Variablen,  $\boldsymbol{\beta} := (\beta_0, \dots, \beta_{k-1})'$  der  $(k \times 1)$ -Vektor der unbekanntenen Regressionskoeffizienten und  $\boldsymbol{\epsilon} := (\epsilon_1, \dots, \epsilon_N)'$  der  $(N \times 1)$ -Vektor der stochastischen Störgrößen.

Je nach dem, welche Annahmen an die Modellkomponenten gestellt werden, unterscheidet man das klassische und das verallgemeinerte lineare Regressionsmodell (Toutenburg, 1992, S. 17 ff.). Im Folgenden wird vorrangig das klassische lineare Regressionsmodell betrachtet. Der Schwerpunkt liegt dabei auf der Herleitung von Prognosen und Prognoseintervallen. Dies setzt jedoch zunächst eine optimale Schätzung der unbekanntenen Modellparameter voraus.

### 3.3.1 Das klassische lineare Regressionsmodell

#### Definition 3.9: Das klassische lineare Regressionsmodell

Sei  $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$  ein Regressionsmodell. Im klassischen linearen Regressionsmodell gelten nach Toutenburg (1992, S. 17) folgende Annahmen für die Modellkomponenten:

- [A1] Die Designmatrix  $\mathbf{X}$  ist eine nicht-stochastische  $(N \times k)$ -Matrix mit  $\text{rang}(\mathbf{X}) = k$ . Die Einstellungen der unabhängigen Variablen werden also als fest vorgegeben bzw. kontrolliert angesehen. Die Anzahl der Beobachtungen ist größer als die Anzahl unbekannter Parameter im Modell, d.h. es gilt  $N > k$ .
- [A2] Der Erwartungswert aller Störgrößen ist 0, sodass  $\mathbb{E}(\boldsymbol{\epsilon}) = \mathbf{0} \in \mathbb{R}^N$  gilt.
- [A3] Die Störgrößen sind unabhängig und identisch verteilt mit Varianz  $\sigma^2 > 0$ . Somit gilt  $\text{Var}(\boldsymbol{\epsilon}) = \sigma^2 \mathbf{I}_N$ , wobei  $\mathbf{I}_N$  die  $(N \times N)$ -Einheitsmatrix ist.
- [A4] Die Störgrößen sind normalverteilt, sodass  $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_N)$  gilt.

Viele Ergebnisse lassen sich auch ohne die Verteilungsannahme [A4] herleiten. Zur Konstruktion von Prognoseintervallen ist diese Annahme jedoch unerlässlich. Die im Folgenden hergeleiteten Resultate gelten auch, wenn die Regressoren stochastisch sind. In diesem Fall würde sich nur die Beweisführung ändern.

### 3.3.1.1 Parameterschätzung

Zu Beginn jeder Regressionsanalyse steht die Schätzung des unbekanntes Koeffizientenvektors  $\beta$ . Eine optimale Schätzfunktion für diesen Parametervektor ist gegeben durch den gewöhnlichen Kleinste-Quadrate-Schätzer

$$\hat{\beta}_{\text{OLS}} := (\hat{\beta}_0, \dots, \hat{\beta}_{k-1})' = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y},$$

wobei OLS für *ordinary least squares* steht. Die Optimalität des OLS-Schätzers  $\hat{\beta}_{\text{OLS}}$  bezieht sich sowohl auf seine algebraischen als auch auf seine stochastischen Eigenschaften. Aus algebraischer Sicht ist  $\hat{\beta}_{\text{OLS}}$  die eindeutige Lösung des Minimierungsproblems

$$\hat{\beta}_{\text{OLS}} = \underset{\beta}{\operatorname{argmin}} (\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta).$$

Der Koeffizientenvektor wird demnach so gewählt, dass die Realisationen der Störgrößen in ihrer Gesamtheit möglichst wenig von 0 abweichen (Toutenburg, 1992, S. 22 ff.).

Während der algebraische Ansatz  $\hat{\beta}_{\text{OLS}}$  als Lösung eines Minimierungsproblems betrachtet, in dem  $\beta$  frei wählbar ist, wird der Koeffizientenvektor beim stochastischen Optimierungsansatz als fest angesehen. Das Ziel ist nun, die einzelnen Koeffizienten  $\beta_i$ ,  $i \in \{0, \dots, k-1\}$ , durch Schätzfunktionen mit möglichst kleiner Varianz zu schätzen. Dazu wird ganz allgemein eine Linearform der Parameter

$$\psi(\beta) = \mathbf{a}'\beta, \quad \mathbf{a} \in \mathbb{R}^k,$$

betrachtet. Im Rahmen der linearen Modelle werden ausschließlich Schätzfunktionen  $\hat{\psi}(\beta)$  zugelassen, die linear in den Beobachtungen  $Y_1, \dots, Y_N$  sowie erwartungstreu für die entsprechende Linearform der Parameter sind. Eine lineare erwartungstreue Schätzfunktion  $\hat{\psi}(\beta)$  für  $\psi(\beta)$  muss also folgende Bedingungen erfüllen (Toutenburg, 1992, S. 33 ff.):

1.  $\hat{\psi}(\beta) = \mathbf{c}'\mathbf{Y}$ ,  $\mathbf{c} \in \mathbb{R}^N$
2.  $\mathbb{E}(\hat{\psi}(\beta)) = \psi(\beta) = \mathbf{a}'\beta$

Diejenige Schätzfunktion  $\hat{\psi}(\beta)$ , die in der Klasse der linearen erwartungstreuen Schätzfunktionen für  $\psi(\beta)$  die kleinste Varianz hat, wird als beste lineare unverzerrte Schätzfunktion – kurz BLUE (*best linear unbiased estimator*) – bezeichnet (Toutenburg, 1992, S. 33 ff.). Der nun folgende Satz von Gauß-Markov zeigt, dass der OLS-Schätzer  $\hat{\beta}_{\text{OLS}}$  eine wichtige Rolle bei der Konstruktion Varianz-optimaler Schätzer spielt. Der Beweis ist Caspary und Wichmann (1994, S. 221) entnommen.



**Satz 3.10: Gauß-Markov-Theorem**

Sei  $Y = X\beta + \epsilon$  ein klassisches lineares Regressionsmodell, in dem die Annahmen [A1]-[A3] gelten. Sei  $\psi(\beta) = a'\beta$  eine zu schätzende Linearform der Parameter. Der beste lineare unverzerrte Schätzer für  $\psi(\beta)$  ist gegeben durch

$$\hat{\psi}(\beta) = a'(X'X)^{-1}X'Y = a'\hat{\beta}_{\text{OLS}}.$$

**Beweis:**

Zunächst wird gezeigt, dass  $\hat{\psi}(\beta) = a'(X'X)^{-1}X'Y$  ein erwartungstreuer Schätzer für  $\psi(\beta) = a'\beta$  ist:

$$\begin{aligned} \mathbb{E}\left(a'(X'X)^{-1}X'Y\right) &= \mathbb{E}\left(a'(X'X)^{-1}X'(X\beta + \epsilon)\right) \\ &= \mathbb{E}\left(a'(X'X)^{-1}X'X\beta + a'(X'X)^{-1}X'\epsilon\right) \\ &= a'(X'X)^{-1}X'X\beta + a'(X'X)^{-1}X'\mathbb{E}(\epsilon) \\ &= a'\beta \quad (\text{nach [A2]}) \end{aligned}$$

Sei nun  $\hat{\psi}(\beta) = c'Y$  der beste lineare unverzerrte Schätzer für  $\psi(\beta) = a'\beta$ . Da  $\hat{\psi}(\beta)$  erwartungstreu für  $\psi(\beta)$  ist, muss gelten:

$$\mathbb{E}(c'Y) = \mathbb{E}(c'(X\beta + \epsilon)) \stackrel{[\text{A2}]}{=} c'X\beta = a'\beta$$

Aus dieser Gleichung folgt, dass  $c'X = a'$  sein muss. Die Varianz von  $\hat{\psi}(\beta)$  lässt sich nun folgendermaßen umschreiben:

$$\begin{aligned} \text{Var}(c'Y) &= \text{Var}\left(\left(c'Y - a'(X'X)^{-1}X'Y\right) + a'(X'X)^{-1}X'Y\right) \\ &= \text{Var}\left(c'Y - a'(X'X)^{-1}X'Y\right) + \text{Var}\left(a'(X'X)^{-1}X'Y\right) \\ &\quad + 2\text{Cov}\left(c'Y - a'(X'X)^{-1}X'Y, a'(X'X)^{-1}X'Y\right) \end{aligned}$$

Für den Kovarianzterm in der letzten Gleichung gilt:

$$\begin{aligned} &\text{Cov}\left(c'Y - a'(X'X)^{-1}X'Y, a'(X'X)^{-1}X'Y\right) \\ &= \text{Cov}\left(\left(c' - a'(X'X)^{-1}X'\right)Y, a'(X'X)^{-1}X'Y\right) \\ &= \left(c' - a'(X'X)^{-1}X'\right) \text{Var}(Y)X(X'X)^{-1}a \\ &= \left(c' - a'(X'X)^{-1}X'\right) \text{Var}(X\beta + \epsilon)X(X'X)^{-1}a \\ &= \left(c' - a'(X'X)^{-1}X'\right) \text{Var}(\epsilon)X(X'X)^{-1}a \\ &= \sigma^2 \left(c' - a'(X'X)^{-1}X'\right) X(X'X)^{-1}a \quad (\text{nach [A3]}) \end{aligned}$$

$$\begin{aligned}
&= \sigma^2(\mathbf{c}'\mathbf{X} - \mathbf{a}')(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a} \\
&= \sigma^2(\mathbf{a}' - \mathbf{a}')(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a} \quad (\text{nach Voraussetzung}) \\
&= 0
\end{aligned}$$

Somit reduziert sich die Varianz von  $\hat{\psi}(\boldsymbol{\beta}) = \mathbf{c}'\mathbf{Y}$  zu:

$$\text{Var}(\mathbf{c}'\mathbf{Y}) = \text{Var}\left(\mathbf{c}'\mathbf{Y} - \mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}\right) + \text{Var}\left(\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}\right)$$

Da  $\hat{\psi}(\boldsymbol{\beta})$  minimale Varianz hat, muss gelten, dass

$$\text{Var}\left(\mathbf{c}'\mathbf{Y} - \mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}\right) = 0$$

ist. Daraus folgt, dass  $\mathbf{c}' = \mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$  sein muss. Damit ist  $\hat{\psi}(\boldsymbol{\beta}) = \mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \mathbf{a}'\hat{\boldsymbol{\beta}}_{\text{OLS}}$  der beste lineare unverzernte Schätzer für  $\psi(\boldsymbol{\beta}) = \mathbf{a}'\boldsymbol{\beta}$ .  $\square$

Im vektorwertigen Sinn ist somit der OLS-Schätzer  $\hat{\boldsymbol{\beta}}_{\text{OLS}}$  der beste lineare unverzernte Schätzer für  $\boldsymbol{\beta}$ . Im folgenden Satz werden die Verteilung und die Konsistenz des OLS-Schätzers hergeleitet.

**Satz 3.11: Verteilung und asymptotische Eigenschaften von  $\hat{\boldsymbol{\beta}}_{\text{OLS}}$**

Sei  $\hat{\boldsymbol{\beta}}_{\text{OLS}} := (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$  der OLS-Schätzer eines klassischen linearen Regressionsmodells  $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ , in dem die Annahmen [A1]-[A4] gelten. Dann gilt nach Seber und Lee (2003, S. 42) für die Verteilung von  $\hat{\boldsymbol{\beta}}_{\text{OLS}}$ :

$$\hat{\boldsymbol{\beta}}_{\text{OLS}} \sim \mathcal{N}(\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1})$$

Falls der Grenzwert  $\mathbf{M} := \lim_{N \rightarrow \infty} (N(\mathbf{X}'\mathbf{X})^{-1})$  existiert und positiv definit ist, ist  $\hat{\boldsymbol{\beta}}_{\text{OLS}}$  nach Greene (2008, S. 64 f.) ein konsistenter Schätzer für  $\boldsymbol{\beta}$ , d.h. es gilt

$$\text{plim}_{N \rightarrow \infty}(\hat{\boldsymbol{\beta}}_{\text{OLS}}) = \boldsymbol{\beta}.$$

**Beweis:**

Für den Erwartungswert und die Varianz von  $\hat{\boldsymbol{\beta}}_{\text{OLS}}$  gilt:

$$\begin{aligned}
\mathbb{E}(\hat{\boldsymbol{\beta}}_{\text{OLS}}) &= \mathbb{E}\left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}\right) = \mathbb{E}\left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon})\right) \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbb{E}(\boldsymbol{\epsilon}) \stackrel{[\text{A2}]}{=} \boldsymbol{\beta}
\end{aligned}$$

$$\begin{aligned}
\text{Var}(\hat{\boldsymbol{\beta}}_{\text{OLS}}) &= \text{Var}\left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}\right) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\text{Var}(\mathbf{Y})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\
&\stackrel{[\text{A3}]}{=} \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}
\end{aligned}$$

Für den Beweis der Konsistenz wird auf Greene (2008, S. 64) verwiesen.  $\square$

Da  $\hat{\beta}_{\text{OLS}}$  der BLUE für  $\beta$  ist, ist  $\hat{Y} := X\hat{\beta}_{\text{OLS}}$  die beste lineare unverzerrte Prognose für den Zufallsvektor  $Y$ . Anhand der Prognose können auch die unbeobachtbaren Realisationen der Störgrößen vorhergesagt werden. Der Vektor  $\hat{e} := Y - \hat{Y}$  enthält die Residuen, d.h. die geschätzten Realisationen der Störgrößen. Der Residuenvektor  $\hat{e}$  ist eine Projektion des Beobachtungsvektors  $Y$  in das orthogonale Komplement des Spaltenraumes der Designmatrix  $X$ , was aus der folgenden Umformung hervorgeht (Toutenburg, 1992, S. 42):

$$\hat{e} = Y - \hat{Y} = Y - X\hat{\beta}_{\text{OLS}} = Y - X(X'X)^{-1}X'Y = (I_N - X(X'X)^{-1}X')Y$$

Die Matrix  $P_{S(X)} := X(X'X)^{-1}X'$  ist ein orthogonaler Projektor in den Spaltenraum von  $X$  und besitzt die im folgenden Satz zusammengefassten Eigenschaften.

**Satz 3.12: Eigenschaften des Projektors  $P_{S(X)}$**

Sei  $X$  eine  $(N \times k)$ -Matrix mit  $\text{rang}(X) = k$ . Für die Matrix  $P_{S(X)} := X(X'X)^{-1}X'$  gelten nach Seber und Lee (2003, S. 40) folgende Eigenschaften:

1.  $P_{S(X)}$  und  $(I_N - P_{S(X)})$  sind symmetrisch und idempotent.
2.  $\text{rang}(I_N - P_{S(X)}) = \text{spur}(I_N - P_{S(X)}) = N - k$ .
3.  $P_{S(X)}X = X$

**Beweis:**

1. Symmetrie und Idempotenz

$$\begin{aligned} P'_{S(X)} &= (X(X'X)^{-1}X')' = X(X'X)^{-1}X' = P_{S(X)} \\ (I_N - P_{S(X)})' &= I_N' - P'_{S(X)} = I_N - P_{S(X)} \\ P_{S(X)}P_{S(X)} &= X(X'X)^{-1}X'X(X'X)^{-1}X' = X(X'X)^{-1}X' = P_{S(X)} \\ (I_N - P_{S(X)})(I_N - P_{S(X)}) &= I_N - 2P_{S(X)} + P_{S(X)}P_{S(X)} = I_N - P_{S(X)} \end{aligned}$$

2. Da  $P_{S(X)}$  idempotent ist, gilt nach Satz 3.7 für den Rang dieser Matrix:

$$\begin{aligned} \text{rang}(I_N - P_{S(X)}) &= \text{spur}(I_N - P_{S(X)}) = \text{spur}(I_N) - \text{spur}(P_{S(X)}) \\ &= N - \text{spur}(X(X'X)^{-1}X') = N - \text{spur}(X'X(X'X)^{-1}) \\ &= N - \text{spur}(I_k) = N - k \end{aligned}$$

3.  $P_{S(X)}X = X(X'X)^{-1}X'X = X$   $\square$

Die Residuen bieten sich als Grundlage für eine Schätzung der Störgrößen-Varianz  $\sigma^2$  natürlicherweise an. Im Folgenden wird ein optimaler Schätzer für  $\sigma^2$  vorgestellt.

**Satz 3.13: Schätzer für  $\sigma^2$**

Sei  $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$  ein klassisches lineares Modell, in dem die Annahmen [A1]-[A3] erfüllt sind. Dann ist

$$\hat{\sigma}^2 := \frac{(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{\text{OLS}})'(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{\text{OLS}})}{N - k} = \frac{\hat{\boldsymbol{\epsilon}}'\hat{\boldsymbol{\epsilon}}}{N - k}$$

ein erwartungstreuer Schätzer der Störgrößen-Varianz  $\sigma^2$  (Casparly und Wichmann, 2004, S. 248 ff.). Unter Hinzunahme der Normalverteilungsannahme [A4] gilt für die Verteilung von  $\hat{\sigma}^2$ :

$$\hat{\sigma}^2 \sim \frac{\sigma^2}{N - k} \chi_{N-k}^2$$

Falls der Grenzwert  $\mathbf{M} := \lim_{N \rightarrow \infty} (N(\mathbf{X}'\mathbf{X})^{-1})$  existiert und positiv definit ist, ist  $\hat{\sigma}^2$  nach Greene (2008, S. 67 ff.) auch ein konsistenter Schätzer für  $\sigma^2$ , d.h. es gilt

$$\text{plim}_{N \rightarrow \infty}(\hat{\sigma}^2) = \sigma^2.$$

**Beweis:**

Für den Beweis der Erwartungstreue und der Konsistenz von  $\hat{\sigma}^2$  wird auf Casparly und Wichmann (2004, S. 248 ff.) sowie Greene (2008, S. 67 ff.) verwiesen. An dieser Stelle wird lediglich die Verteilung von  $\hat{\sigma}^2$  hergeleitet:

Sei  $\mathbf{Z} := (\sigma^{-1}\mathbf{Y})$ . Dann gilt  $\mathbf{Z} \sim \mathcal{N}(\sigma^{-1}\mathbf{X}\boldsymbol{\beta}, \mathbf{I}_N)$ , weil  $\mathbb{E}(\mathbf{Z}) = \sigma^{-1}\mathbb{E}(\mathbf{Y}) = \sigma^{-1}\mathbf{X}\boldsymbol{\beta}$  und  $\text{Var}(\mathbf{Z}) = \sigma^{-2}\text{Var}(\mathbf{Y}) = \sigma^{-2}\sigma^2\mathbf{I}_N = \mathbf{I}_N$  ist. Die Residuenquadratsumme kann nun unter Ausnutzung des Resultats von Satz 3.12 (1) wie folgt umgeschrieben werden:

$$\hat{\boldsymbol{\epsilon}}'\hat{\boldsymbol{\epsilon}} = \mathbf{Y}'(\mathbf{I}_N - \mathbf{P}_{S(\mathbf{X})})'(\mathbf{I}_N - \mathbf{P}_{S(\mathbf{X})})\mathbf{Y} = \mathbf{Y}'(\mathbf{I}_N - \mathbf{P}_{S(\mathbf{X})})\mathbf{Y} = \sigma^2\mathbf{Z}'(\mathbf{I}_N - \mathbf{P}_{S(\mathbf{X})})\mathbf{Z}$$

Da die Kovarianzmatrix von  $\mathbf{Z}$  die Einheitsmatrix ist und weil  $(\mathbf{I}_N - \mathbf{P}_{S(\mathbf{X})})$  eine idempotente, symmetrische Matrix mit Rang  $(N - k)$  ist, gilt nach Satz 3.8, dass  $\hat{\boldsymbol{\epsilon}}'\hat{\boldsymbol{\epsilon}} \sim \sigma^2 \chi_{N-k, \delta}^2$  verteilt ist mit Nichtzentralitätsparameter

$$\delta = (\sigma^{-1}\mathbf{X}\boldsymbol{\beta})'(\mathbf{I}_N - \mathbf{P}_{S(\mathbf{X})})(\sigma^{-1}\mathbf{X}\boldsymbol{\beta}) \stackrel{\text{Satz 3.12 (3)}}{=} \sigma^{-2}\boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta} - \sigma^{-2}\boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = 0.$$

Für die Verteilung von  $\hat{\sigma}^2$  gilt also:  $\hat{\sigma}^2 = \frac{\hat{\boldsymbol{\epsilon}}'\hat{\boldsymbol{\epsilon}}}{N - k} \sim \frac{\sigma^2}{N - k} \chi_{N-k}^2$ .

□

### 3.3.1.2 Punkt- und Intervallprognosen für zukünftige Beobachtungen

Eine zentrale Funktion linearer Modelle besteht in der Prognose einer zukünftigen Beobachtung  $Y_0 = \mathbf{x}_0' \boldsymbol{\beta} + \epsilon_0$ . Hierbei ist  $\mathbf{x}_0 := (1, x_{01}, \dots, x_{0(k-1)})'$  der Vektor, der die Einstellungen der unabhängigen Variablen enthält. Da  $\widehat{\boldsymbol{\beta}}_{\text{OLS}}$  der BLUE für  $\boldsymbol{\beta}$  ist, ist  $\widehat{Y}_0 := \mathbf{x}_0' \widehat{\boldsymbol{\beta}}_{\text{OLS}}$  die beste lineare unverzerrte Prognose für die Beobachtung  $Y_0$  (Seber und Lee, 2003, S. 131 ff.).

Neben der Punktprognose kann die Information des bereits beobachteten Zufallsvektors  $\mathbf{Y}$  auch zur Konstruktion von Intervallprognosen verwendet werden. Für eine Intervallprognose werden die Ergebnisse einer früheren Stichprobe genutzt, um eine neue Beobachtung mit vorgegebener Wahrscheinlichkeit zu überdecken. Im folgenden Satz wird ein  $(1 - \alpha)$ -Prognoseintervall für die Beobachtung  $Y_0$  im klassischen linearen Modell hergeleitet. Zunächst wird das Prognoseintervall jedoch formal definiert.

**Definition 3.14: Prognoseintervall im klassischen linearen Modell**

Sei  $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$  ein klassisches lineares Regressionsmodell und sei  $Y_0 = \mathbf{x}_0' \boldsymbol{\beta} + \epsilon_0$  eine neue, noch unbekannte Beobachtung. Ein Intervall  $\widehat{\mathcal{PI}}(\mathbf{Y}) := [u, o]$  ( $u, o \in \mathbb{R}$  mit  $u \leq o$ ), welches die neue Beobachtung  $Y_0$  mit Wahrscheinlichkeit  $(1 - \alpha)$  überdeckt, wird als Prognoseintervall bezeichnet. Ein Prognoseintervall  $\widehat{\mathcal{PI}}(\mathbf{Y})$  hängt stets von dem bereits beobachteten Zufallsvektor  $\mathbf{Y} = (Y_1, \dots, Y_N)'$  ab. Es gilt

$$P_{(\boldsymbol{\beta}, \sigma^2)} \left( Y_0 \in \widehat{\mathcal{PI}}(\mathbf{Y}) \right) = 1 - \alpha \quad \forall \boldsymbol{\beta} \in \mathbb{R}^k, \forall \sigma^2 \in \mathbb{R}.$$

Ein Prognoseintervall unterscheidet sich wesentlich von einem Konfidenzintervall, welches Linearkombinationen der festen Parameter mit vorgegebener Wahrscheinlichkeit überdeckt. Prognoseintervalle sind stets breiter als Konfidenzintervalle, da es im Vergleich zum Konfidenzintervall eine zusätzliche Varianzquelle gibt: Zunächst müssen – wie beim Konfidenzintervall auch – die unbekannt Parameter auf Basis der alten Stichprobe geschätzt werden. Bei einem Prognoseintervall kommt jedoch noch die Varianz der neuen Beobachtung hinzu. Da nur die Varianz der Parameterschätzer durch den Stichprobenumfang kontrolliert werden kann, konvergiert ein Prognoseintervall im Gegensatz zu einem Konfidenzintervall für  $N \rightarrow \infty$  immer gegen ein Intervall und nicht gegen einen festen Wert (Patel, 1989).

**Satz 3.15: Konstruktion eines Prognoseintervalls für eine zukünftige Beobachtung**

Sei  $Y = X\beta + \epsilon$  ein klassisches lineares Regressionsmodell, für welches die Annahmen [A1]-[A4] erfüllt sind. Sei  $Y_0 = x_0'\beta + \epsilon_0$  eine neue Beobachtung, für die  $\epsilon_0 \sim \mathcal{N}(0, \sigma^2)$  gilt. Ferner seien  $\epsilon_0, \epsilon_1, \dots, \epsilon_N$  stochastisch unabhängig. Dann gilt nach Seber und Lee (2003, S. 131 f.):

$$\frac{\hat{Y}_0 - Y_0}{\sqrt{\hat{v}_p}} \sim t_{N-k},$$

wobei die geschätzte Prognosevarianz gegeben ist durch  $\hat{v}_p := \hat{\sigma}^2 (1 + x_0'(X'X)^{-1}x_0)$ . Somit lautet das exakte  $(1 - \alpha)$ -Prognoseintervall für  $Y_0$ :

$$\widehat{\mathcal{PI}}(Y) = \left[ \hat{Y}_0 - t_{N-k, 1-\alpha/2} \sqrt{\hat{v}_p}, \hat{Y}_0 + t_{N-k, 1-\alpha/2} \sqrt{\hat{v}_p} \right]$$

Hierbei bezeichnet  $t_{N-k, 1-\alpha/2}$  das  $(1 - \alpha/2)$ -Quantil der  $t_{N-k}$ -Verteilung.

**Beweis:**

Zunächst wird die Verteilung des Prognosefehlers  $(\hat{Y}_0 - Y_0)$  betrachtet:

Es gilt  $(\hat{Y}_0 - Y_0) \sim \mathcal{N}(0, v_p)$ , wobei  $v_p := \sigma^2 (1 + x_0'(X'X)^{-1}x_0)$  die Varianz des Prognosefehlers ist:

$$\begin{aligned} \mathbb{E}(\hat{Y}_0 - Y_0) &= \mathbb{E}(x_0'\hat{\beta}_{\text{OLS}} - x_0'\beta - \epsilon_0) = x_0'\mathbb{E}(\hat{\beta}_{\text{OLS}}) - x_0'\beta - \mathbb{E}(\epsilon_0) \\ &= x_0'\beta - x_0'\beta = 0 \end{aligned}$$

Da nach Voraussetzung  $\epsilon_0, \epsilon_1, \dots, \epsilon_N$  stochastisch unabhängig sind, ist die Punktprognose  $\hat{Y}_0$  stochastisch unabhängig von der neuen Beobachtung  $Y_0$ . Die Varianz des Prognosefehlers ist daher gerade die Summe der beiden Komponenten:

$$\begin{aligned} \text{Var}(\hat{Y}_0 - Y_0) &= \text{Var}(x_0'\hat{\beta}_{\text{OLS}} - x_0'\beta - \epsilon_0) \stackrel{\epsilon_0, \epsilon_1, \dots, \epsilon_N \text{ unabhg.}}{=} \text{Var}(x_0'\hat{\beta}_{\text{OLS}}) + \text{Var}(\epsilon_0) \\ &\stackrel{\text{Satz 3.11}}{=} \sigma^2 x_0'(X'X)^{-1}x_0 + \sigma^2 = \sigma^2 (1 + x_0'(X'X)^{-1}x_0) \end{aligned}$$

Nun wird der durch die geschätzte Prognosevarianz standardisierte Prognosefehler in Komponenten zerlegt, deren Verteilungen bekannt sind. Es werden die beiden Zufallsvariablen  $Z := \frac{\hat{Y}_0 - Y_0}{\sqrt{\hat{v}_p}}$  und  $W := \frac{N-k}{\sigma^2} \hat{\sigma}^2$  definiert. Dann gilt:

$$\frac{Z}{\sqrt{\frac{1}{N-k}W}} = \frac{\hat{Y}_0 - Y_0}{\sqrt{\hat{v}_p}}$$

Hinsichtlich der Verteilung von  $Z$  und  $W$  gilt:  $Z \sim \mathcal{N}(0, 1)$  und  $W \sim \chi_{N-k}^2$ , da

- $\mathbb{E}(Z) = \frac{\mathbb{E}(\hat{Y}_0 - Y_0)}{\sqrt{\hat{v}_p}} = 0$

- $\text{Var}(Z) = \frac{\text{Var}(\hat{Y}_0 - Y_0)}{v_p} = 1$
- Aus Satz 3.13 folgt für die Verteilung von  $W$ :  

$$W = \frac{N-k}{\sigma^2} \hat{\sigma}^2 \sim \frac{N-k}{\sigma^2} \frac{\sigma^2}{N-k} \chi_{N-k}^2 = \chi_{N-k}^2$$

Zuletzt muss noch gezeigt werden, dass  $Z$  und  $W$  stochastisch unabhängig sind:

$Y_0$  ist stochastisch unabhängig von  $\hat{\sigma}^2$ , da  $Y_0$  stochastisch unabhängig von  $\mathbf{Y} = (Y_1, \dots, Y_N)'$  ist. Auch  $\hat{Y}_0 = \mathbf{x}_0' \hat{\boldsymbol{\beta}}_{\text{OLS}}$  ist stochastisch unabhängig von  $\hat{\sigma}^2$ , da die Linearformen  $\hat{\boldsymbol{\beta}}_{\text{OLS}}$  und  $\hat{\boldsymbol{\varepsilon}} = (\mathbf{I}_N - \mathbf{P}_{S(\mathbf{X})})\mathbf{Y}$  stochastisch unabhängig sind. Aufgrund der Normalverteilungsannahme genügt es nach Satz 3.2 zu zeigen, dass  $\hat{\boldsymbol{\beta}}_{\text{OLS}}$  und  $\hat{\boldsymbol{\varepsilon}}$  unkorreliert sind:

$$\begin{aligned} \text{Cov}(\hat{\boldsymbol{\beta}}_{\text{OLS}}, \hat{\boldsymbol{\varepsilon}}) &= \text{Cov}\left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}, (\mathbf{I}_N - \mathbf{P}_{S(\mathbf{X})})\mathbf{Y}\right) \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\text{Var}(\mathbf{Y})(\mathbf{I}_N - \mathbf{P}_{S(\mathbf{X})}) \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{I}_N - \mathbf{P}_{S(\mathbf{X})}) \quad (\text{nach [A3]}) \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' - \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \\ &= \mathbf{0} \end{aligned}$$

Da der Prognosefehler  $(\hat{Y}_0 - Y_0)$  und der Varianzschätzer  $\hat{\sigma}^2$  stochastisch unabhängig sind, sind auch die Zufallsvariablen  $Z$  und  $W$  stochastisch unabhängig. Nach der Definition der  $t$ -Verteilung (3.4) gilt demnach:

$$\frac{\hat{Y}_0 - Y_0}{\sqrt{\hat{v}_p}} \sim t_{N-k}$$

Das  $(1 - \alpha)$ -Prognoseintervall für  $Y_0$  ergibt sich dann aus der folgenden Umformung:

$$\begin{aligned} &\mathbb{P}_{(\boldsymbol{\beta}, \sigma^2)} \left( t_{N-k, \alpha/2} \leq \frac{\hat{Y}_0 - Y_0}{\sqrt{\hat{v}_p}} \leq t_{N-k, 1-\alpha/2} \right) = 1 - \alpha \\ \Leftrightarrow &\mathbb{P}_{(\boldsymbol{\beta}, \sigma^2)} \left( \hat{Y}_0 - t_{N-k, 1-\alpha/2} \sqrt{\hat{v}_p} \leq Y_0 \leq \hat{Y}_0 + t_{N-k, 1-\alpha/2} \sqrt{\hat{v}_p} \right) = 1 - \alpha \end{aligned}$$

□

### 3.3.2 Das verallgemeinerte lineare Regressionsmodell

Der Unterschied zwischen dem klassischen und dem verallgemeinerten linearen Regressionsmodell (GLM) besteht ausschließlich in der Annahme [A3], welche die Kovarianzstruktur der Störgrößen betrifft. Im verallgemeinerten linearen Regressionsmodell wird die restriktive Annahme [A3] dahin gehend abgeschwächt, dass die Kovarianzmatrix der Störgrößen lediglich positiv definit sein muss. Die modifizierte Annahme lautet (Toutenburg, 1992, S. 18 ff.):

[A3\*] Die Kovarianzmatrix der Störgrößen lautet  $\text{Var}(\epsilon) = \sigma^2 \mathbf{\Omega}$ . Dabei ist  $\sigma^2 > 0$  und  $\mathbf{\Omega} \in \mathbb{R}^{N \times N}$  eine positiv definite Matrix.

Diese abgeschwächte Annahme lässt also sowohl stochastische Abhängigkeit zwischen den Störgrößen (Autokorrelation) als auch unterschiedliche Varianzen der Störgrößen (Heteroskedastizität) zu (Toutenburg, 1992, S. 18 ff.). Im folgenden Satz wird der beste lineare erwartungstreue Schätzer für den Koeffizientenvektor  $\beta$  im verallgemeinerten linearen Regressionsmodell angegeben. Dieser wird später zur optimalen Parameterschätzung im Regressionsmodell mit zufälligen Effekten (Abschnitt 3.4) benötigt.

#### Satz 3.16: Beste lineare unverzerrte Parameterschätzung im GLM

Sei  $Y = X\beta + \epsilon$  ein verallgemeinertes lineares Regressionsmodell, für das die Annahmen [A1], [A2] und [A3\*] gelten. Dann ist der beste lineare unverzerrte Schätzer für  $\beta$  nach Toutenburg (1992, S. 103) gegeben durch den Aitken-Schätzer

$$\hat{\beta}_{\text{GLS}} := (\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{Y}.$$

Die Kurzform GLS steht hierbei für *generalized least squares*.

Der Parametervektor kann im GLM also nur dann optimal geschätzt werden, wenn die Kovarianzstruktur der Störgrößen (d.h. wenn  $\mathbf{\Omega}$ ) bekannt ist.



### 3.4 Das Regressionsmodell mit zufälligen Effekten (RCR-Modell)

Im Gegensatz zum klassischen linearen Regressionsmodell ermöglicht das Regressionsmodell mit zufälligen Effekten, alle Beobachtungsreihen gleichzeitig zu modellieren. Auf diese Weise können Prognosen für einen neuen Verlauf erstellt werden, die auch die Informationen bereits vergangener Verläufe berücksichtigen. Im Folgenden wird dieses Regressionsmodell zunächst allgemein vorgestellt, um später verschiedene Prognoseszenarien zu betrachten.

Bei longitudinalen Daten können die einzelnen Verläufe häufig gut durch ein Polynom der Zeit beschrieben werden. Der funktionale Zusammenhang zwischen der Zeit und der abhängigen Variablen von Verlauf  $i$  ist dann gegeben durch

$$Y_{ij} = \beta_{0i} + \beta_{1i}x_{1ij} + \beta_{2i}x_{2ij} + \dots + \beta_{(k-1)i}x_{(k-1)ij} + \epsilon_{ij}, \quad j \in \{1, \dots, N_i\},$$

wobei der verlaufsspezifische Parametervektor durch  $\beta_i := (\beta_{0i}, \dots, \beta_{(k-1)i})'$  gegeben ist,  $x_{ij} := (1, x_{1ij}, \dots, x_{(k-1)ij})'$  den Vektor der zeitlichen Polynomialterme darstellt (z.B.  $x_{1ij} = t_{ij}$ ,  $x_{2ij} = t_{ij}^2, \dots$ ) und  $\epsilon_{ij}$  eine stochastische Störgröße mit Erwartungswert 0 ist (Swamy, 1971, S. 1 ff.). In Matrixform kann das Modell für alle Verläufe wie folgt dargestellt werden:

$$Y_i = X_i \beta_i + \epsilon_i, \quad i = 1 \dots, I.$$

Hierbei ist  $Y_i := (Y_{i1}, \dots, Y_{iN_i})'$  der Beobachtungsvektor der abhängigen Variablen von Verlauf  $i$ ,  $X_i := (x_{i1}, \dots, x_{iN_i})'$  die  $(N_i \times k)$ -Designmatrix der unabhängigen Variablen und  $\epsilon_i := (\epsilon_{i1}, \dots, \epsilon_{iN_i})'$  der stochastische Störvektor mit Erwartungswert  $\mathbf{0}$ . Im Rahmen dieser Bachelorarbeit werden balancierte Daten mit gleichen Zeitpunkten analysiert. Die Designmatrizen der einzelnen Verläufe sind also identisch. Im Folgenden wird die Designmatrix daher unabhängig vom Verlauf mit  $X$  bezeichnet.

Ein Analyseansatz bestünde darin, für jeden Verlauf ein individuelles klassisches lineares Regressionsmodell anzupassen. Das Ziel der longitudinalen Datenanalyse ist jedoch, die Verläufe aller experimentellen Einheiten durch ein einziges Modell zu erfassen. Die verschiedenen Beobachtungsreihen könnten nur dann durch ein einheitliches klassisches Regressionsmodell modelliert werden, wenn sich alle Verläufe über die Zeit hinweg gleich entwickeln würden, d.h. wenn sich Verlauf  $i$  wie Verlauf  $j$  verhält. Dies ist jedoch nicht der Fall: Das Problem besteht nämlich darin, dass die Koeffizientenvektoren  $\beta_i$ ,  $i \in \{1 \dots, I\}$ , verlaufsspezifisch sind. Das heißt, dass jeder Verlauf seinen eigenen Achsenabschnitt, seine eigene Steigung usw. aufweist. Die Variabilität der Koeffizientenvektoren ist darauf zurückzuführen, dass nur die Zeit als unabhängige Variable zur Verfügung steht und wichtige (unbekannte) Variablen, die den Verlauf

einer Einheit näher spezifizieren könnten, im Modell fehlen (Swamy, 1971, S. 2).

Um die interindividuelle Parametervariation mit zu modellieren, werden die Koeffizientenvektoren  $\beta_1, \dots, \beta_I$  als unabhängige Realisationen einer gemeinsamen multivariaten Verteilung, d.h. als zufällig, angesehen (Gumpertz und Pantula, 1989). In Matrixform lässt sich das Regressionsmodell mit zufälligen Effekten für den Fall balancierter Daten mit gleichen Zeitpunkten formulieren als

$$Y_i = X\beta_i + \epsilon_i, \quad i = 1, \dots, I.$$

In der Literatur ist das Regressionsmodell mit zufälligen Effekten unter dem Begriff RCR-Modell (*random coefficient regression model*) verbreitet. Das hier vorgestellte RCR-Modell ist das einfachste seiner Art, da für alle Verläufe dieselben Beobachtungszeitpunkte gelten. Es wurde von Rao bereits 1965 aufgestellt (Rao, 1965). Die von Rao entwickelte Schätz- und Testtheorie für dieses Modell wurde von Swamy 1971 auf den Fall balancierter Daten mit verschiedenen Beobachtungszeitpunkten und letztendlich von Carter, Yang und Johansen in den 1980er Jahren auf den Fall unbalancierter Daten erweitert (Swamy, 1971; Carter und Yang, 1986; Johansen, 1982). Im Folgenden wird das RCR-Modell für balancierte, longitudinale Daten mit gleichen Zeitpunkten formal definiert.

### Definition 3.17: Das Regressionsmodell mit zufälligen Effekten

Ein RCR-Modell für balancierte Longitudinaldaten mit gleichen Zeitpunkten besteht aus  $I$  separaten Regressionsgleichungen der Form  $Y_i = X\beta_i + \epsilon_i$ ,  $i \in \{1, \dots, I\}$ . Dabei werden folgende Annahmen für die Modellkomponenten gemacht (Swamy, 1971, S. 98 ff.):

- [R1]  $X$  ist eine nicht-stochastische  $(N \times k)$ -Designmatrix, welche die Einstellungen der unabhängigen Variablen (Polynomterme der Zeitpunkte) enthält. Die Anzahl der Messzeitpunkte ist größer als die Anzahl der zu schätzenden Parameter, d.h. es gilt  $N > k$ . Zudem hat  $X$  vollen Rang, d.h. es gilt  $\text{rang}(X) = k$ .
- [R2] Die stochastischen Koeffizientenvektoren  $\beta_i$ ,  $i \in \{1, \dots, I\}$ , sind unabhängig und identisch verteilt mit Erwartungswert  $b \in \mathbb{R}^k$  und regulärer Kovarianzmatrix  $\Sigma \in \mathbb{R}^{k \times k}$ . Der stochastische Charakter der Koeffizientenvektoren kann anders ausgedrückt werden als  $\beta_i = b + \delta_i$ , wobei  $\delta_i := (\delta_{0i}, \dots, \delta_{(k-1)i})'$  ein Zufallsvektor mit Erwartungswert  $\mathbf{0}$  und Kovarianzmatrix  $\Sigma$  ist. Die Parameter in  $b := (b_0, \dots, b_{k-1})'$  werden als „feste (nicht-stochastische) Effekte“ bezeichnet. Der Vektor  $b$  ist der „mittlere“ Parametervektor der Gesamtpopulation. Die Realisationen von  $\delta_i$  werden als „zufällige Effekte“ bezeichnet. Somit ist jeder Koeffizientenvektor  $\beta_i$  als Summe von globalen, festen und verlaufsspezifischen, zufälligen Effekten zu interpretieren.

- [R3] Die stochastischen  $(N \times 1)$ -Störvektoren  $\epsilon_i$ ,  $i \in \{1, \dots, I\}$ , sind unabhängig und identisch verteilt mit Erwartungswert  $\mathbf{0}$  und Kovarianzmatrix  $\sigma^2 \mathbf{I}_N$ , wobei  $\sigma^2 > 0$  ist und  $\mathbf{I}_N$  die  $(N \times N)$ -Einheitsmatrix bezeichnet.
- [R4] Die Zufallsvektoren  $\beta_i$  und  $\epsilon_j$  sind stochastisch unabhängig für alle  $i, j \in \{1, \dots, I\}$ , d.h.  $\beta_1, \dots, \beta_I, \epsilon_1, \dots, \epsilon_I$  sind stochastisch unabhängig.
- [R5] Sowohl die Koeffizientenvektoren  $\beta_i$ ,  $i \in \{1, \dots, I\}$ , als auch die Störvektoren  $\epsilon_i$ ,  $i \in \{1, \dots, I\}$ , sind multivariat normalverteilt. Damit gilt  $\beta_i \sim \mathcal{N}(\mathbf{b}, \Sigma)$  und  $\epsilon_i \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_N)$  für alle  $i \in \{1, \dots, I\}$ .

Die Normalverteilungsannahme [R5] ist für die Herleitung optimaler Parameterschätzer nicht notwendig. Sie wird jedoch für die spätere Konstruktion von Prognoseintervallen benötigt. Der zentrale Unterschied zum klassischen linearen Regressionsmodell besteht also darin, dass im RCR-Modell die Koeffizienten zufällig sind. Der stochastische Charakter der Koeffizienten ist darauf zurückzuführen, dass die experimentellen Einheiten des Datensatzes als Zufallsstichprobe aus einer großen Population betrachtet werden, in der jede Einheit (d.h. jeder Verlauf) seinen eigenen Koeffizientenvektor besitzt. Für eine gegebene Einheit aus dieser Population (z.B. für einen bestimmten Dialysetag) ist der Parametervektor  $\beta_i$  dann aber für alle sukzessiven Beobachtungen dieser Einheit gleich (Bondeson, 1990).

Im RCR-Modell gibt es zwei Quellen der Varianz: Zum einen wird die interindividuelle Heterogenität der Verläufe durch die zufälligen Koeffizienten modelliert. Zum anderen werden die intraindividuellen Messfehler durch die stochastischen Störgrößen berücksichtigt. Diese beiden Variabilitätskomponenten erlauben es, die gesamten  $I$  Verläufe durch ein gemeinsames Regressionsmodell zu beschreiben (Verbeke und Molenberghs, 2000, S. 2).

**Satz 3.18: Verteilung der verlaufsspezifischen Beobachtungsvektoren  $Y_1, \dots, Y_I$**

Sei  $Y_i = X\beta_i + \epsilon_i$ ,  $i \in \{1, \dots, I\}$ , ein RCR-Modell, für das die Annahmen [R1]-[R5] gelten. Aus den Annahmen des RCR-Modells folgt, dass die Beobachtungsvektoren  $Y_1, \dots, Y_I$  unabhängig und identisch normalverteilt sind mit Erwartungswert  $X\mathbf{b}$  und Kovarianzmatrix  $(X\Sigma X' + \sigma^2 \mathbf{I}_N)$ , d.h.

$$Y_1, \dots, Y_I \stackrel{u.i.v.}{\sim} \mathcal{N}(X\mathbf{b}, X\Sigma X' + \sigma^2 \mathbf{I}_N).$$

**Beweis:**

Aus Annahme [R4] folgt, dass  $Y_1, \dots, Y_I$  stochastisch unabhängig sind. Für den Erwartungswert von  $Y_i$  gilt:

$$\mathbb{E}(Y_i) = \mathbb{E}(X\beta_i + \epsilon_i) = X\mathbb{E}(\beta_i) + \mathbb{E}(\epsilon_i) \stackrel{[\text{R2}], [\text{R3}]}{=} X\mathbf{b}$$

Für die Varianz von  $Y_i$  gilt:

$$\text{Var}(Y_i) = \text{Var}(X\beta_i + \epsilon_i) \stackrel{[\text{R4}]}{=} X\text{Var}(\beta_i)X' + \text{Var}(\epsilon_i) \stackrel{[\text{R2}], [\text{R3}]}{=} X\Sigma X' + \sigma^2 I_N \quad \square$$

Somit sind alle Verläufe unabhängige Realisationen derselben multivariaten Normalverteilung. Während die Beobachtungen zweier verschiedener Verläufe stochastisch unabhängig voneinander sind, sind die Beobachtungen  $Y_{i1}, \dots, Y_{iN}$  eines Verlaufs – in der Regel positiv – korreliert (Frees, 2006, S. 7). Bedingt man die Verteilung von  $Y_i$  allerdings auf eine feste Realisation von  $\beta_i$ , so ergibt sich ein klassisches lineares Regressionsmodell  $Y_i = X\beta_i + \epsilon_i$ , in dem  $(Y_i|\beta_i) \sim \mathcal{N}(X\beta_i, \sigma^2 I_N)$  gilt (Weiss, 2005, S. 304 ff.). Die unbedingte Verteilung von  $Y_i$  wird auch als Marginalverteilung bezeichnet (Laird und Ware, 1982).

Wie aus den Erläuterungen zu Annahme [R2] hervorgeht, lässt sich jeder Koeffizientenvektor  $\beta_i$  als Summe der festen Effekte  $\mathbf{b}$  und der zufälligen Effekte  $\delta_i$  darstellen. Die  $I$  Regressionsgleichungen können daher folgendermaßen umgeformt werden (Poi, 2003):

$$\begin{aligned} Y_i &= X\beta_i + \epsilon_i, \quad i = 1, \dots, I \\ \Leftrightarrow Y_i &= X(\mathbf{b} + \delta_i) + \epsilon_i, \quad i = 1, \dots, I \\ \Leftrightarrow Y_i &= X\mathbf{b} + X\delta_i + \epsilon_i, \quad i = 1, \dots, I \end{aligned}$$

Das RCR-Modell ist ein Spezialfall eines linearen gemischten Modells. Diese breite Klasse von Regressionsmodellen umfasst alle Regressionsmodelle, die sowohl feste als auch zufällige Effekte enthalten. Während im RCR-Modell alle Koeffizienten als zufällig angesehen werden, kann in linearen gemischten Modellen auch nur eine Teilmenge der  $k$  Koeffizienten als stochastisch spezifiziert werden (Frees, 2006, S. 86 ff.).

### 3.4.1 Parameterschätzung im RCR-Modell

In diesem Abschnitt werden Schätzer für die unbekannt, festen Parameter des RCR-Modells hergeleitet. Dabei handelt es sich um die  $k$  festen Effekte des Parametervektors  $\mathbf{b}$ , um die  $\frac{k(k+1)}{2}$  Komponenten der Kovarianzmatrix  $\Sigma$  sowie um die intraindividuelle Varianz  $\sigma^2$ . Die Anzahl unbekannter Parameter ist somit größer als im klassischen linearen Regressionsmodell. Es werden auch asymptotische Verteilungseigenschaften der Schätzer betrachtet, da diese für die Herleitung approximativer Prognoseintervalle nötig sind.

### 3.4.1.1 Schätzung der festen Effekte

Für die Herleitung des besten linearen unverzerrten Schätzers des Parametervektors  $\mathbf{b}$  werden die  $I$  Regressionsgleichungen  $\mathbf{Y}_i = \mathbf{X}\boldsymbol{\beta}_i + \boldsymbol{\epsilon}_i$ ,  $i \in \{1, \dots, I\}$ , zunächst getrennt betrachtet. Sei dazu  $\hat{\boldsymbol{\beta}}_{i,\text{OLS}} := (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}_i$  der OLS-Schätzer des verlaufsspezifischen Parametervektors  $\boldsymbol{\beta}_i$ . Dessen Verteilung wird basierend auf Gumpertz und Pantula (1989) im nächsten Satz hergeleitet.

#### Satz 3.19: Verteilung des OLS-Schätzers $\hat{\boldsymbol{\beta}}_{i,\text{OLS}}$ im RCR-Modell

Sei  $\mathbf{Y}_i = \mathbf{X}\boldsymbol{\beta}_i + \boldsymbol{\epsilon}_i$ ,  $i \in \{1, \dots, I\}$ , ein RCR-Modell, für welches die Annahmen [R1]-[R5] gelten. Dann sind die verlaufsspezifischen OLS-Schätzer  $\hat{\boldsymbol{\beta}}_{1,\text{OLS}}, \dots, \hat{\boldsymbol{\beta}}_{I,\text{OLS}}$  unabhängig und identisch normalverteilt mit Erwartungswert  $\mathbf{b}$  und Kovarianzmatrix  $(\boldsymbol{\Sigma} + \sigma^2(\mathbf{X}'\mathbf{X})^{-1})$ , d.h.

$$\hat{\boldsymbol{\beta}}_{1,\text{OLS}}, \dots, \hat{\boldsymbol{\beta}}_{I,\text{OLS}} \stackrel{u.i.v.}{\sim} \mathcal{N}(\mathbf{b}, \boldsymbol{\Sigma} + \sigma^2(\mathbf{X}'\mathbf{X})^{-1})$$

Die Varianz von  $\hat{\boldsymbol{\beta}}_{i,\text{OLS}}$  ist also die Summe aus der unkontrollierbaren Populationsvarianz  $\boldsymbol{\Sigma}$  und der kontrollierbaren Designvarianz  $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$ .

#### Beweis:

Die stochastische Unabhängigkeit von  $\hat{\boldsymbol{\beta}}_{1,\text{OLS}}, \dots, \hat{\boldsymbol{\beta}}_{I,\text{OLS}}$  folgt direkt aus der Unabhängigkeit von  $\mathbf{Y}_1, \dots, \mathbf{Y}_I$  (siehe Satz 3.18). Für den Erwartungswert und für die Varianz von  $\hat{\boldsymbol{\beta}}_{i,\text{OLS}}$ ,  $i \in \{1, \dots, I\}$ , gilt:

$$\begin{aligned} \mathbb{E}(\hat{\boldsymbol{\beta}}_{i,\text{OLS}}) &= \mathbb{E}\left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}_i\right) = \mathbb{E}\left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta}_i + \boldsymbol{\epsilon}_i)\right) \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\mathbb{E}(\boldsymbol{\beta}_i) + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbb{E}(\boldsymbol{\epsilon}_i) \stackrel{[\text{R2}], [\text{R3}]}{=} \mathbf{b} \end{aligned}$$

$$\begin{aligned} \text{Var}(\hat{\boldsymbol{\beta}}_{i,\text{OLS}}) &= \text{Var}\left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}_i\right) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\text{Var}(\mathbf{Y}_i)\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\Sigma}\mathbf{X}' + \sigma^2\mathbf{I}_N)\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \quad (\text{nach Satz 3.18}) \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\Sigma}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} + \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &= \boldsymbol{\Sigma} + \sigma^2(\mathbf{X}'\mathbf{X})^{-1} \end{aligned}$$

□

Der einfachste denkbare Schätzer für den Parametervektor  $\mathbf{b}$  ist das arithmetische Mittel der verlaufsspezifischen OLS-Schätzer  $\hat{\boldsymbol{\beta}}_{i,\text{OLS}}$ ,  $i \in \{1, \dots, I\}$ . Im nächsten Satz wird gezeigt, dass dieser Schätzer auch tatsächlich der beste lineare erwartungstreue Schätzer für  $\mathbf{b}$  ist. Für den Beweis wird eine spezielle Matrixumformung benötigt, welche im folgenden Lemma vorgestellt wird.

**Lemma 3.20: Inverse von  $(\mathbf{X}\boldsymbol{\Sigma}\mathbf{X}' + \sigma^2\mathbf{I}_N)$** 

Sei  $\sigma^2 > 0$ ,  $\mathbf{I}_N$  die  $(N \times N)$ -Einheitsmatrix,  $\boldsymbol{\Sigma} \in \mathbb{R}^{k \times k}$  nichtsingulär und  $\mathbf{X}$  eine  $(N \times k)$ -Matrix,  $N > k$ , mit  $\text{rang}(\mathbf{X}) = k$ . Dann ist die Inverse von  $(\mathbf{X}\boldsymbol{\Sigma}\mathbf{X}' + \sigma^2\mathbf{I}_N)$  nach Swamy (1971, S. 101) gegeben durch

$$\sigma^{-2}\mathbf{I}_N - \sigma^{-2}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \left( \boldsymbol{\Sigma} + \sigma^2(\mathbf{X}'\mathbf{X})^{-1} \right)^{-1} (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' .$$

**Beweis:**

Durch die Linksmultiplikation mit  $(\mathbf{X}\boldsymbol{\Sigma}\mathbf{X}' + \sigma^2\mathbf{I}_N)$  ergibt sich die  $(N \times N)$ -Einheitsmatrix:

$$\begin{aligned} & (\mathbf{X}\boldsymbol{\Sigma}\mathbf{X}' + \sigma^2\mathbf{I}_N) \left( \sigma^{-2}\mathbf{I}_N - \sigma^{-2}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \left( \boldsymbol{\Sigma} + \sigma^2(\mathbf{X}'\mathbf{X})^{-1} \right)^{-1} (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \right) \\ &= \sigma^{-2}\mathbf{X}\boldsymbol{\Sigma}\mathbf{X}' - \sigma^{-2}\mathbf{X}\boldsymbol{\Sigma}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{X}\boldsymbol{\Sigma}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \left( \boldsymbol{\Sigma} + \sigma^2(\mathbf{X}'\mathbf{X})^{-1} \right)^{-1} (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \\ &\quad + \mathbf{I}_N - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \sigma^2\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \left( \boldsymbol{\Sigma} + \sigma^2(\mathbf{X}'\mathbf{X})^{-1} \right)^{-1} (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \\ &= \sigma^{-2}\mathbf{X}\boldsymbol{\Sigma}\mathbf{X}' - \sigma^{-2}\mathbf{X}\boldsymbol{\Sigma}\mathbf{X}' + \mathbf{X}\boldsymbol{\Sigma} \left( \boldsymbol{\Sigma} + \sigma^2(\mathbf{X}'\mathbf{X})^{-1} \right)^{-1} (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \\ &\quad + \mathbf{I}_N - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \sigma^2\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \left( \boldsymbol{\Sigma} + \sigma^2(\mathbf{X}'\mathbf{X})^{-1} \right)^{-1} (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \\ &= \mathbf{I}_N - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \left( \mathbf{X}\boldsymbol{\Sigma} + \sigma^2\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \right) \left( \boldsymbol{\Sigma} + \sigma^2(\mathbf{X}'\mathbf{X})^{-1} \right)^{-1} (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \\ &= \mathbf{I}_N - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{X} \left( \boldsymbol{\Sigma} + \sigma^2(\mathbf{X}'\mathbf{X})^{-1} \right) \left( \boldsymbol{\Sigma} + \sigma^2(\mathbf{X}'\mathbf{X})^{-1} \right)^{-1} (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \\ &= \mathbf{I}_N - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \\ &= \mathbf{I}_N \end{aligned} \quad \square$$

**Satz 3.21: Bester linearer erwartungstreuer Schätzer der festen Effekte**

Sei  $\mathbf{Y}_i = \mathbf{X}\boldsymbol{\beta}_i + \boldsymbol{\epsilon}_i$ ,  $i \in \{1, \dots, I\}$ , ein RCR-Modell, für das die Annahmen [R1]-[R4] gelten. Dann ist

$$\hat{\mathbf{b}}_{\text{GLS}} := \frac{1}{I} \sum_{i=1}^I \hat{\boldsymbol{\beta}}_{i,\text{OLS}}$$

der beste lineare unverzerrte Schätzer für den Parametervektor  $\mathbf{b}$  (Swamy, 1971, S. 99 ff.). Unter Hinzunahme der Normalverteilungsannahme [R5] gilt für die Verteilung dieses Schätzers (Swamy, 1971, S. 18 ff.):

$$\hat{\mathbf{b}}_{\text{GLS}} \sim \mathcal{N} \left( \mathbf{b}, \frac{1}{I} \left( \boldsymbol{\Sigma} + \sigma^2(\mathbf{X}'\mathbf{X})^{-1} \right) \right).$$

**Beweis:**

Zunächst wird das Modell zu einer kompakteren Form umformuliert. Anstatt die  $I$  Regressionsgleichungen isoliert zu betrachten, können die Gleichungen aller Einhei-

ten zu einer gepoolten Regressionsgleichung zusammengefasst werden:

Sei  $\mathbf{Y} := (\mathbf{Y}_1', \dots, \mathbf{Y}_I')$  der  $(NI \times 1)$ -Vektor, der die Beobachtungen aller Verläufe enthält. Sei  $\mathcal{X} := [\mathbf{X}', \dots, \mathbf{X}']'$  die  $(NI \times k)$ -Matrix, die aus  $I$  untereinander gesetzten Designmatrizen  $\mathbf{X}$  besteht. Desweiteren sei  $\mathbf{D}(\mathbf{X}) = \mathbf{I}_I \otimes \mathbf{X}$  die  $(NI \times Ik)$ -Matrix, auf deren  $I$  diagonalen  $(N \times k)$ -Blöcken die Designmatrix  $\mathbf{X}$  und auf deren sonstigen Blöcken eine Null-Matrix steht. Außerdem werden die zufälligen Effekte aller Verläufe in dem  $(Ik \times 1)$ -Vektor  $\boldsymbol{\delta} := (\boldsymbol{\delta}_1', \dots, \boldsymbol{\delta}_I')$  und die Störvektoren aller Individuen in dem  $(NI \times 1)$ -Vektor  $\boldsymbol{\epsilon} := (\boldsymbol{\epsilon}_1', \dots, \boldsymbol{\epsilon}_I')$  zusammengefasst. Mit dieser Notation lassen sich die  $I$  Verläufe durch die folgende Gleichung zusammenfassen (Swamy, 1971, S. 99 ff.):

$$\begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \vdots \\ \mathbf{Y}_I \end{bmatrix} = \begin{bmatrix} \mathbf{X} \\ \mathbf{X} \\ \vdots \\ \mathbf{X} \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_{k-1} \end{bmatrix} + \begin{bmatrix} \mathbf{X} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{X} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{X} \end{bmatrix} \begin{bmatrix} \boldsymbol{\delta}_1 \\ \boldsymbol{\delta}_2 \\ \vdots \\ \boldsymbol{\delta}_I \end{bmatrix} + \begin{bmatrix} \boldsymbol{\epsilon}_1 \\ \boldsymbol{\epsilon}_2 \\ \vdots \\ \boldsymbol{\epsilon}_I \end{bmatrix}$$

$$\Leftrightarrow \mathbf{Y} = \mathcal{X}\mathbf{b} + \mathbf{D}(\mathbf{X})\boldsymbol{\delta} + \boldsymbol{\epsilon}$$

Aus Annahme [R4] folgt, dass  $\boldsymbol{\delta}$  und  $\boldsymbol{\epsilon}$  stochastisch unabhängig sind. Aus den Annahmen [R2], [R3] und [R5] folgt mit Satz 3.18:  $\boldsymbol{\delta} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_I \otimes \boldsymbol{\Sigma})$  und  $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_{NI})$  (Swamy, 1971, S. 99 ff.).

Sei nun  $\mathbf{u} := \mathbf{D}(\mathbf{X})\boldsymbol{\delta} + \boldsymbol{\epsilon}$ . Für den Erwartungswert des stochastischen Vektors  $\mathbf{u}$  gilt:

$$\mathbb{E}(\mathbf{u}) = \mathbb{E}(\mathbf{D}(\mathbf{X})\boldsymbol{\delta} + \boldsymbol{\epsilon}) = \mathbf{D}(\mathbf{X})\mathbb{E}(\boldsymbol{\delta}) + \mathbb{E}(\boldsymbol{\epsilon}) = \mathbf{0}$$

Da der Erwartungswert des Vektors  $\mathbf{u}$  gerade  $\mathbf{0}$  ist, kann  $\mathbf{u}$  als Störvektor des verallgemeinerten linearen Regressionsmodells

$$\mathbf{Y} = \mathcal{X}\mathbf{b} + \mathbf{u}$$

aufgefasst werden. Die Kovarianzmatrix  $\mathbf{H} := \text{Var}(\mathbf{u})$  ist gegeben durch:

$$\begin{aligned} & \text{Var}(\mathbf{D}(\mathbf{X})\boldsymbol{\delta} + \boldsymbol{\epsilon}) \stackrel{\boldsymbol{\delta}, \boldsymbol{\epsilon} \text{ unabhg.}}{=} \text{Var}(\mathbf{D}(\mathbf{X})\boldsymbol{\delta}) + \text{Var}(\boldsymbol{\epsilon}) \\ &= \mathbf{D}(\mathbf{X})\text{Var}(\boldsymbol{\delta})\mathbf{D}(\mathbf{X})' + \text{Var}(\boldsymbol{\epsilon}) \\ &= (\mathbf{I}_I \otimes \mathbf{X})(\mathbf{I}_I \otimes \boldsymbol{\Sigma})(\mathbf{I}_I \otimes \mathbf{X})' + \sigma^2 \mathbf{I}_{NI} \\ &= \begin{bmatrix} \mathbf{X} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{X} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{X} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \boldsymbol{\Sigma} \end{bmatrix} \begin{bmatrix} \mathbf{X} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{X} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{X} \end{bmatrix}' + \sigma^2 \begin{bmatrix} \mathbf{I}_N & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_N & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{I}_N \end{bmatrix} \end{aligned}$$

$$\begin{aligned}
&= \begin{bmatrix} (\mathbf{X}\boldsymbol{\Sigma}\mathbf{X}' + \sigma^2\mathbf{I}_N) & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & (\mathbf{X}\boldsymbol{\Sigma}\mathbf{X}' + \sigma^2\mathbf{I}_N) & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & (\mathbf{X}\boldsymbol{\Sigma}\mathbf{X}' + \sigma^2\mathbf{I}_N) \end{bmatrix} \\
&= \mathbf{I}_I \otimes (\mathbf{X}\boldsymbol{\Sigma}\mathbf{X}' + \sigma^2\mathbf{I}_N)
\end{aligned}$$

Daraus ergibt sich, dass das RCR-Modell letztlich ein verallgemeinertes lineares Modell mit spezieller Kovarianzstruktur ist. Mithilfe des Aitken-Schätzers aus Satz 3.16 kann nun der beste lineare unverzerrte Schätzer für  $\mathbf{b}$  hergeleitet werden. Der folgende Beweis stammt in modifizierter Form aus einem Artikel von Kadiyala und Oberhelman (1982). Bei der mit (\*) markierten Gleichheit wird das Resultat aus Lemma 3.20 angewendet:

$$\begin{aligned}
\hat{\mathbf{b}}_{\text{GLS}} &= (\mathbf{X}'\mathbf{H}^{-1}\mathbf{X})^{-1} \mathbf{X}'\mathbf{H}^{-1}\mathbf{Y} \\
&= \left( \sum_{i=1}^I \mathbf{X}' (\mathbf{X}\boldsymbol{\Sigma}\mathbf{X}' + \sigma^2\mathbf{I}_N)^{-1} \mathbf{X} \right)^{-1} \sum_{i=1}^I \mathbf{X}' (\mathbf{X}\boldsymbol{\Sigma}\mathbf{X}' + \sigma^2\mathbf{I}_N)^{-1} \mathbf{Y}_i \\
&= \left( \mathbf{I}\mathbf{X}' (\mathbf{X}\boldsymbol{\Sigma}\mathbf{X}' + \sigma^2\mathbf{I}_N)^{-1} \mathbf{X} \right)^{-1} \mathbf{X}' (\mathbf{X}\boldsymbol{\Sigma}\mathbf{X}' + \sigma^2\mathbf{I}_N)^{-1} \sum_{i=1}^I \mathbf{Y}_i \\
&\stackrel{*}{=} \frac{1}{I} \left[ \mathbf{X}' \left( \sigma^{-2}\mathbf{I}_N - \sigma^{-2}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} (\boldsymbol{\Sigma} + \sigma^2(\mathbf{X}'\mathbf{X})^{-1})^{-1} (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \right) \mathbf{X} \right]^{-1} \\
&\quad \times \mathbf{X}' \left( \sigma^{-2}\mathbf{I}_N - \sigma^{-2}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} (\boldsymbol{\Sigma} + \sigma^2(\mathbf{X}'\mathbf{X})^{-1})^{-1} (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \right) \sum_{i=1}^I \mathbf{Y}_i \\
&= \frac{1}{I} \left[ \sigma^{-2}\mathbf{X}'\mathbf{X} - \sigma^{-2}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X} + \mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} (\boldsymbol{\Sigma} + \sigma^2(\mathbf{X}'\mathbf{X})^{-1})^{-1} (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X} \right]^{-1} \\
&\quad \times \left( \sigma^{-2}\mathbf{X}' - \sigma^{-2}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} (\boldsymbol{\Sigma} + \sigma^2(\mathbf{X}'\mathbf{X})^{-1})^{-1} (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \right) \sum_{i=1}^I \mathbf{Y}_i \\
&= \frac{1}{I} \left[ (\boldsymbol{\Sigma} + \sigma^2(\mathbf{X}'\mathbf{X})^{-1})^{-1} \right]^{-1} (\boldsymbol{\Sigma} + \sigma^2(\mathbf{X}'\mathbf{X})^{-1})^{-1} (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \sum_{i=1}^I \mathbf{Y}_i \\
&= \frac{1}{I} (\boldsymbol{\Sigma} + \sigma^2(\mathbf{X}'\mathbf{X})^{-1}) (\boldsymbol{\Sigma} + \sigma^2(\mathbf{X}'\mathbf{X})^{-1})^{-1} (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \sum_{i=1}^I \mathbf{Y}_i \\
&= \frac{1}{I} \sum_{i=1}^I (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}_i \\
&= \frac{1}{I} \sum_{i=1}^I \hat{\boldsymbol{\beta}}_{i,\text{OLS}}
\end{aligned}$$



Hinsichtlich des Erwartungswertes von  $\widehat{\mathbf{b}}_{\text{GLS}}$  gilt:

$$\mathbb{E}(\widehat{\mathbf{b}}_{\text{GLS}}) = \mathbb{E}\left(\frac{1}{I} \sum_{i=1}^I \widehat{\boldsymbol{\beta}}_{i,\text{OLS}}\right) = \frac{1}{I} \sum_{i=1}^I \mathbb{E}(\widehat{\boldsymbol{\beta}}_{i,\text{OLS}}) = \frac{1}{I} I\mathbf{b} = \mathbf{b}$$

Da  $\widehat{\boldsymbol{\beta}}_{1,\text{OLS}}, \dots, \widehat{\boldsymbol{\beta}}_{I,\text{OLS}}$  nach Satz 3.19 stochastisch unabhängig sind, ist die Varianz von  $\widehat{\mathbf{b}}_{\text{GLS}}$  gerade die Summe der Varianzen von  $\widehat{\boldsymbol{\beta}}_{i,\text{OLS}}, i \in \{1, \dots, I\}$ :

$$\begin{aligned} \text{Var}(\widehat{\mathbf{b}}_{\text{GLS}}) &= \text{Var}\left(\frac{1}{I} \sum_{i=1}^I \widehat{\boldsymbol{\beta}}_{i,\text{OLS}}\right) \\ &= \frac{1}{I^2} \sum_{i=1}^I \text{Var}(\widehat{\boldsymbol{\beta}}_{i,\text{OLS}}) \quad (\text{da } \widehat{\boldsymbol{\beta}}_{1,\text{OLS}}, \dots, \widehat{\boldsymbol{\beta}}_{I,\text{OLS}} \text{ unabhängig}) \\ &= \frac{1}{I^2} I(\boldsymbol{\Sigma} + \sigma^2(\mathbf{X}'\mathbf{X})^{-1}) \quad (\text{nach Satz 3.19}) \\ &= \frac{1}{I}(\boldsymbol{\Sigma} + \sigma^2(\mathbf{X}'\mathbf{X})^{-1}) \end{aligned}$$

□

Es wurde mit dem letzten Satz gezeigt, dass der beste lineare unverzerrte Schätzer für  $\mathbf{b}$  bei balancierten Daten mit gleichen Zeitpunkten unabhängig von den Varianzparametern  $\boldsymbol{\Sigma}$  und  $\sigma^2$  ist. Dies ist bei unbalancierten Designs nicht der Fall: Hier hängt  $\widehat{\mathbf{b}}_{\text{GLS}}$  von den Varianzparametern ab, sodass diese für eine optimale Schätzung von  $\mathbf{b}$  bekannt sein müssen (Swamy, 1971, S. 99 ff.).

### 3.4.1.2 Schätzung der Varianzparameter

Wie bereits gezeigt, kommt die optimale Schätzung der festen Effekte ohne die Kenntnis der Varianzkomponenten aus. Zur Herleitung von Prognoseintervallen werden jedoch Schätzer für die Kovarianzmatrix der zufälligen Effekte  $\boldsymbol{\Sigma}$  und für die Störgrößenvarianz  $\sigma^2$  benötigt. Für die Konsistenz der im Folgenden vorgestellten Schätzer ist die Annahme von Nöten, dass der Grenzwert  $\mathbf{M} := \lim_{N \rightarrow \infty} (N(\mathbf{X}'\mathbf{X})^{-1})$  existiert und positiv definit ist. Dies ist auf jeden Fall erfüllt, wenn die Anzahl der verschiedenen Zeitpunkte gegen  $\infty$  strebt (Gumpertz und Pantula, 1989).

#### Satz 3.22: Schätzer für $\sigma^2$

Sei  $\mathbf{Y}_i = \mathbf{X}\boldsymbol{\beta}_i + \boldsymbol{\epsilon}_i, i \in \{1, \dots, I\}$ , ein RCR-Modell, für das die Annahmen [R1]-[R4] gelten. Ein erwartungstreuer, konsistenter Schätzer für die intraindividuelle Fehlervarianz  $\sigma^2$  ist nach Swamy (1971, S. 106 ff.) gegeben durch das arithmetische Mittel der

Varianzschätzer aus den  $I$  Regressionsgleichungen, d.h. durch

$$\hat{\sigma}_{\text{RCR}}^2 := \frac{1}{I} \sum_{i=1}^I \hat{\sigma}_i^2 = \frac{1}{I} \sum_{i=1}^I \frac{(\mathbf{Y}_i - \mathbf{X}\hat{\boldsymbol{\beta}}_{i,\text{OLS}})'(\mathbf{Y}_i - \mathbf{X}\hat{\boldsymbol{\beta}}_{i,\text{OLS}})}{N-k}.$$

Hierbei bezeichnet  $\hat{\sigma}_i^2$ ,  $i \in \{1, \dots, I\}$ , die gewöhnliche Varianzschätzung des  $i$ -ten Regressionsmodells  $\mathbf{Y}_i = \mathbf{X}\boldsymbol{\beta}_i + \boldsymbol{\epsilon}_i$ .

**Beweis:**

Aus der Theorie des klassischen linearen Regressionsmodells folgt, dass  $\hat{\sigma}_i^2$  für alle  $i \in \{1, \dots, I\}$  ein erwartungstreuer Schätzer für  $\sigma^2$  ist. Damit gilt:

$$\mathbb{E} \left( \hat{\sigma}_{\text{RCR}}^2 \right) = \mathbb{E} \left( \frac{1}{I} \sum_{i=1}^I \hat{\sigma}_i^2 \right) = \frac{1}{I} \sum_{i=1}^I \mathbb{E} \left( \hat{\sigma}_i^2 \right) = \sigma^2$$

Falls der Grenzwert  $\mathbf{M} = \lim_{N \rightarrow \infty} (N(\mathbf{X}'\mathbf{X})^{-1})$  existiert, ist  $\hat{\sigma}_i^2$  nach Satz 3.13 für alle  $i \in \{1, \dots, I\}$  ein konsistenter Schätzer für  $\sigma^2$ . Somit gilt auch die Konsistenz von  $\hat{\sigma}_{\text{RCR}}^2$ :

$$\text{plim}_{N \rightarrow \infty} \left( \hat{\sigma}_{\text{RCR}}^2 \right) = \text{plim}_{N \rightarrow \infty} \left( \frac{1}{I} \sum_{i=1}^I \hat{\sigma}_i^2 \right) = \frac{1}{I} \sum_{i=1}^I \text{plim}_{N \rightarrow \infty} \left( \hat{\sigma}_i^2 \right) = \sigma^2 \quad \square$$

Für die Konsistenz von  $\hat{\sigma}_{\text{RCR}}^2$  genügt es also, dass die Anzahl der Beobachtungen pro Verlauf gegen  $\infty$  strebt. Die Anzahl der Verläufe kann beliebig sein.

**Satz 3.23: Schätzer für  $\boldsymbol{\Sigma}$**

Sei  $\mathbf{Y}_i = \mathbf{X}\boldsymbol{\beta}_i + \boldsymbol{\epsilon}_i$ ,  $i \in \{1, \dots, I\}$ , ein RCR-Modell, für das die Annahmen [R1]-[R4] gelten. Ein erwartungstreuer, konsistenter Schätzer für die Kovarianzmatrix der zufälligen Effekte  $\boldsymbol{\Sigma}$  ist nach Swamy (1971, S. 106 ff) gegeben durch

$$\hat{\boldsymbol{\Sigma}} := \frac{1}{I-1} \sum_{i=1}^I \left( \hat{\boldsymbol{\beta}}_{i,\text{OLS}} - \hat{\mathbf{b}}_{\text{GLS}} \right) \left( \hat{\boldsymbol{\beta}}_{i,\text{OLS}} - \hat{\mathbf{b}}_{\text{GLS}} \right)' - \hat{\sigma}_{\text{RCR}}^2 (\mathbf{X}'\mathbf{X})^{-1}.$$

**Beweis:**

Zunächst wird die Erwartungstreue von  $\hat{\boldsymbol{\Sigma}}$  gezeigt:

$$\begin{aligned} & \mathbb{E} \left( \frac{1}{I-1} \sum_{i=1}^I \left( \hat{\boldsymbol{\beta}}_{i,\text{OLS}} - \hat{\mathbf{b}}_{\text{GLS}} \right) \left( \hat{\boldsymbol{\beta}}_{i,\text{OLS}} - \hat{\mathbf{b}}_{\text{GLS}} \right)' - \hat{\sigma}_{\text{RCR}}^2 (\mathbf{X}'\mathbf{X})^{-1} \right) \\ &= \frac{1}{I-1} \sum_{i=1}^I \mathbb{E} \left( \left( \hat{\boldsymbol{\beta}}_{i,\text{OLS}} - \mathbf{b} + \mathbf{b} - \hat{\mathbf{b}}_{\text{GLS}} \right) \left( \hat{\boldsymbol{\beta}}_{i,\text{OLS}} - \mathbf{b} + \mathbf{b} - \hat{\mathbf{b}}_{\text{GLS}} \right)' \right) - \mathbb{E} \left( \hat{\sigma}_{\text{RCR}}^2 (\mathbf{X}'\mathbf{X})^{-1} \right) \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{I-1} \sum_{i=1}^I \mathbb{E} \left( \left( \hat{\beta}_{i,\text{OLS}} - \mathbf{b} \right) \left( \hat{\beta}_{i,\text{OLS}} - \mathbf{b} \right)' + \left( \hat{\beta}_{i,\text{OLS}} - \mathbf{b} \right) \left( \mathbf{b} - \hat{\mathbf{b}}_{\text{GLS}} \right)' \right. \\
&\quad \left. + \left( \mathbf{b} - \hat{\mathbf{b}}_{\text{GLS}} \right) \left( \hat{\beta}_{i,\text{OLS}} - \mathbf{b} \right)' + \left( \mathbf{b} - \hat{\mathbf{b}}_{\text{GLS}} \right) \left( \mathbf{b} - \hat{\mathbf{b}}_{\text{GLS}} \right)' \right) - \mathbb{E} \left( \hat{\sigma}_{\text{RCR}}^2 (\mathbf{X}'\mathbf{X})^{-1} \right) \\
&= \frac{1}{I-1} \left[ \sum_{i=1}^I \mathbb{E} \left( \hat{\beta}_{i,\text{OLS}} - \mathbf{b} \right) \left( \hat{\beta}_{i,\text{OLS}} - \mathbf{b} \right)' - I \mathbb{E} \left( \left( \hat{\mathbf{b}}_{\text{GLS}} - \mathbf{b} \right) \left( \hat{\mathbf{b}}_{\text{GLS}} - \mathbf{b} \right)' \right) \right. \\
&\quad \left. - I \mathbb{E} \left( \left( \hat{\mathbf{b}}_{\text{GLS}} - \mathbf{b} \right) \left( \hat{\mathbf{b}}_{\text{GLS}} - \mathbf{b} \right)' \right) + I \mathbb{E} \left( \left( \hat{\mathbf{b}}_{\text{GLS}} - \mathbf{b} \right) \left( \hat{\mathbf{b}}_{\text{GLS}} - \mathbf{b} \right)' \right) \right] - \mathbb{E} \left( \hat{\sigma}_{\text{RCR}}^2 (\mathbf{X}'\mathbf{X})^{-1} \right) \\
&= \frac{1}{I-1} \left( I \text{Var} \left( \hat{\beta}_{i,\text{OLS}} \right) - I \text{Var} \left( \hat{\mathbf{b}}_{\text{GLS}} \right) \right) - \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} \\
&= \frac{1}{I-1} \left( I \left( \boldsymbol{\Sigma} + \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} \right) - I \frac{1}{I} \left( \boldsymbol{\Sigma} + \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} \right) \right) - \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} \quad (\text{Satz 3.19 und 3.21}) \\
&= \frac{1}{I-1} \left( (I-1) \left( \boldsymbol{\Sigma} + \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} \right) \right) - \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} \\
&= \boldsymbol{\Sigma} + \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} - \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} \\
&= \boldsymbol{\Sigma}
\end{aligned}$$

Um die Konsistenz des Schätzers  $\hat{\boldsymbol{\Sigma}}$  zu zeigen, muss sowohl die Anzahl der Beobachtungen pro Verlauf ( $N$ ) als auch die Anzahl der Verläufe ( $I$ ) gegen  $\infty$  streben. Der im Folgenden skizzierte Beweis ist so aufgebaut, dass zunächst  $N$  und in einem zweiten Schritt  $I$  gegen  $\infty$  strebt (Swamy, 1971, S. 116 ff.). Die Konsistenz gilt also nur dann, wenn  $N$  wesentlich schneller gegen  $\infty$  strebt als  $I$  (Johansen, 1982).

Aus der Konsistenz von  $\hat{\sigma}_{\text{RCR}}^2$  für  $N \rightarrow \infty$  und aus der Annahme, dass der Grenzwert  $\mathbf{M} := \lim_{N \rightarrow \infty} (N(\mathbf{X}'\mathbf{X})^{-1})$  existiert und positiv definit ist, folgt dieses Zwischenergebnis (ZE 1):

$$\begin{aligned}
\text{plim}_{N \rightarrow \infty} \left( \hat{\sigma}_{\text{RCR}}^2 (\mathbf{X}'\mathbf{X})^{-1} \right) &= \text{plim}_{N \rightarrow \infty} \left( \hat{\sigma}_{\text{RCR}}^2 \right) \lim_{N \rightarrow \infty} \left( \frac{1}{N} \right) \lim_{N \rightarrow \infty} \left( N(\mathbf{X}'\mathbf{X})^{-1} \right) \\
&= \sigma^2 \cdot 0 \cdot \mathbf{M} \\
&= \mathbf{0}
\end{aligned} \tag{ZE 1}$$

Swamy hat gezeigt, dass die Matrix  $\hat{\mathbf{S}} := \sum_{i=1}^I \left( \hat{\beta}_{i,\text{OLS}} - \hat{\mathbf{b}}_{\text{GLS}} \right) \left( \hat{\beta}_{i,\text{OLS}} - \hat{\mathbf{b}}_{\text{GLS}} \right)'$  für  $N \rightarrow \infty$  in Wahrscheinlichkeit gegen die Matrix  $\mathbf{S} := \sum_{i=1}^I \left( \beta_i - \bar{\beta} \right) \left( \beta_i - \bar{\beta} \right)'$  konvergiert (siehe Swamy, 1971, S. 116 ff.):

$$\text{plim}_{N \rightarrow \infty} \left( \sum_{i=1}^I \left( \hat{\beta}_{i,\text{OLS}} - \hat{\mathbf{b}}_{\text{GLS}} \right) \left( \hat{\beta}_{i,\text{OLS}} - \hat{\mathbf{b}}_{\text{GLS}} \right)' \right) = \sum_{i=1}^I \left( \beta_i - \bar{\beta} \right) \left( \beta_i - \bar{\beta} \right)' \tag{ZE 2}$$

Hierbei ist  $\bar{\beta} := \frac{1}{I} \sum_{i=1}^I \beta_i$  das arithmetische Mittel der wahren, verlaufsspezifischen

Koeffizientenvektoren. Insgesamt ergibt sich daraus folgendes Konvergenzverhalten von  $\widehat{\Sigma}$  für  $N \rightarrow \infty$ :

$$\begin{aligned} \text{plim}_{N \rightarrow \infty} \widehat{\Sigma} &= \text{plim}_{N \rightarrow \infty} \left( \frac{\widehat{S}}{I-1} - \widehat{\sigma}_{\text{RCR}}^2 (\mathbf{X}'\mathbf{X})^{-1} \right) \\ &= \text{plim}_{N \rightarrow \infty} \left( \frac{\widehat{S}}{I-1} \right) - \text{plim}_{N \rightarrow \infty} \left( \widehat{\sigma}_{\text{RCR}}^2 (\mathbf{X}'\mathbf{X})^{-1} \right) \\ &= \frac{S}{I-1} \quad (\text{nach ZE 1 und ZE 2}) \end{aligned}$$

Also konvergiert  $\widehat{\Sigma}$  für  $N \rightarrow \infty$  in Wahrscheinlichkeit gegen die wahre empirische Kovarianzmatrix der unbekanntem Koeffizientenvektor-Stichprobe  $\{\beta_1, \dots, \beta_I\}$ .

Die empirische Kovarianzmatrix ist immer ein konsistenter Schätzer der wahren Kovarianzmatrix. Wenn man nun im zweiten Schritt  $I$  gegen  $\infty$  streben lässt, gilt also:

$$\text{plim}_{I \rightarrow \infty} \frac{S}{I-1} = \Sigma$$

Zusammenfassend lässt sich der Konvergenzprozess des Schätzers  $\Sigma$  folgendermaßen zweistufig darstellen:

$$\widehat{\Sigma} \xrightarrow{N \rightarrow \infty} \frac{S}{I-1} \xrightarrow{I \rightarrow \infty} \Sigma \quad \square$$

Es wurde gezeigt, dass sowohl  $\widehat{\sigma}_{\text{RCR}}^2$  als auch  $\widehat{\Sigma}$  unverzerrte und – unter bestimmten Annahmen an den Konvergenzprozess – auch konsistente Schätzer für  $\sigma^2$  bzw.  $\Sigma$  sind. Ein fundamentales Problem des Schätzers  $\Sigma$  besteht darin, dass er nicht notwendigerweise positiv definit ist. Dies ist eine störende Eigenschaft des Schätzers, da negative Varianzkomponenten in der Praxis nicht interpretierbar sind. Daher gibt es zahlreiche Modifikationen des Schätzers  $\widehat{\Sigma}$ , die dessen positive Definitheit garantieren (siehe z.B. Carter und Yang, 1986).

Heutzutage werden die von Swamy eingeführten Schätzer  $\widehat{\sigma}_{\text{RCR}}^2$  und  $\widehat{\Sigma}$  nur noch selten verwendet. Stattdessen werden die Varianzkomponenten mithilfe des Maximum-Likelihood- oder des Restricted-Maximum-Likelihood-Verfahrens über rechenintensive, iterative Algorithmen geschätzt (Frees, 2006, S. 100 ff.).

### 3.4.2 Prädiktion der zufälligen Effekte

Mit der Schätzung des globalen Parametervektors  $\mathbf{b}$  kann prognostiziert werden, wie sich die Gesamtpopulation zu zukünftigen Zeitpunkten verhalten wird. Bei vielen Anwendungen besteht jedoch das Interesse darin, zukünftige Beobachtung einzelner Verläufe vorherzusagen (Bondeson, 1990). Angenommen, das Interesse bestünde in der Vorhersage einer zukünftigen Beobachtung von Verlauf  $i$ ,  $i \in \{1, \dots, I\}$ , zum Zeitpunkt  $T$ . Sei  $\mathbf{x}_T := (1, T, T^2, \dots)'$  der  $(k \times 1)$ -Vektor der unabhängigen Variablen, d.h. der Polynomterme von  $T$ . Die vorherzusagende Beobachtung ist dann gegeben durch

$$Y_{iT} = \mathbf{x}_T' \boldsymbol{\beta}_i + \epsilon_{iT} = \mathbf{x}_T' \mathbf{b} + \mathbf{x}_T' \boldsymbol{\delta}_i + \epsilon_{iT}.$$

Das Ziel ist also, die Linearkombination  $\mathbf{x}_T' \mathbf{b} + \mathbf{x}_T' \boldsymbol{\delta}_i$  aus festen und zufälligen Effekten zu schätzen (Bondeson, 1990). Der Begriff „Schätzen“ ist jedoch an dieser Stelle unangebracht, da  $\boldsymbol{\beta}_i$  im RCR-Modell kein fester Parametervektor, sondern ein Zufallsvektor ist. Es handelt sich also vielmehr um eine Vorhersage der Realisation von  $\boldsymbol{\beta}_i$ . Um diese Problematik von der klassischen Schätztheorie abzugrenzen, werden optimale „Schätzfunktionen“ für Linearkombinationen aus festen und zufälligen Effekten als BLUP's (beste lineare unverzerrte Prädiktoren) bezeichnet (Robinson, 1991). Im Folgenden werden die Eigenschaften eines BLUP's näher definiert.

**Definition 3.24: Linearer unverzerrter Prädiktor**

Sei  $\widehat{W}$  ein linearer unverzerrter Prädiktor für die Linearkombination  $W := \boldsymbol{\lambda}' \mathbf{b} + \boldsymbol{\lambda}' \boldsymbol{\delta}_i$ , wobei  $\boldsymbol{\lambda}$  ein fester, frei wählbarer  $(k \times 1)$ -Vektor ist. Für den Prädiktor  $\widehat{W}$  gelten folgende Eigenschaften (Harville, 1976):

1. Sei  $\mathbf{Y} := (\mathbf{Y}_1', \dots, \mathbf{Y}_I')'$  der Vektor, der die Beobachtungen von allen  $I$  Verläufen enthält. Ein Prädiktor  $\widehat{W}$  ist linear, falls er von der Form  $\widehat{W} = \mathbf{r}' \mathbf{Y}$  ist. Hierbei ist  $\mathbf{r}$  ein fester  $(NI \times 1)$ -Vektor.
2. Ein linearer Prädiktor  $\widehat{W}$  für  $W$  ist unverzerrt, wenn der Erwartungswert des Prädiktors  $\widehat{W}$  dem Erwartungswert der vorherzusagenden Größe  $W$  entspricht, d.h. wenn gilt

$$\mathbb{E}(\widehat{W}) = \mathbb{E}(W) = \mathbb{E}(\boldsymbol{\lambda}' \mathbf{b} + \boldsymbol{\lambda}' \boldsymbol{\delta}_i) = \boldsymbol{\lambda}' \mathbf{b}.$$

Da die zu „schätzende“ Linearkombination  $W$  selbst wieder eine Zufallsvariable ist, gilt nicht wie bei gewöhnlichen Schätzern  $\mathbb{E}(\widehat{W}) = W$ .

Während bei Schätzern für feste Parameter die Minimierung der Varianz des Schätzers als Optimalitätskriterium verwendet wird, betrachtet man bei Prädiktoren  $\widehat{W}$  den mittleren quadratischen Fehler, der im Folgenden nach Harville (1976) definiert wird.

**Definition 3.25: Mittlerer quadratischer Fehler**

Sei  $\widehat{W} = r'Y$ ,  $r \in \mathbb{R}^{NI}$ , ein linearer Prädiktor für die Linearkombination  $W := \lambda'b + \lambda'\delta_i$ ,  $\lambda \in \mathbb{R}^k$ . Dann ist der mittlere quadratische Fehler des Prädiktors  $\widehat{W}$  gegeben durch

$$\text{MSE}(\widehat{W}, W) = \mathbb{E} \left( (\widehat{W} - W)^2 \right).$$

Mithilfe dieser Definitionen kann nun der beste lineare unverzerrte Prädiktor für  $W$  wie folgt formal definiert werden (Harville, 1976):

**Definition 3.26: Bester linearer unverzerrter Prädiktor (BLUP)**

Sei  $\mathcal{W} := \{ \widehat{W} \mid \widehat{W} \text{ linear und unverzerrt für } W \}$  die Menge aller linearen und unverzerrten Prädiktoren für die Linearkombination  $W := \lambda'b + \lambda'\delta_i$ ,  $\lambda \in \mathbb{R}^k$ . Der Prädiktor  $\widehat{W}_{\text{BLUP}} \in \mathcal{W}$  heißt bester linearer unverzerrter Prädiktor für  $W$ , falls  $\widehat{W}_{\text{BLUP}}$  den kleinsten mittleren quadratischen Fehler in der Klasse der linearen unverzerrten Prädiktoren hat. Es gilt also

$$\text{MSE} \left( \widehat{W}_{\text{BLUP}}, W \right) \leq \text{MSE} \left( \widehat{W}, W \right) \quad \forall \widehat{W} \in \mathcal{W}.$$

Die Herleitung geschlossener Formeln für BLUP's geht auf Harville (1976) zurück, der das Gauss-Markov-Theorem für die optimale „Vorhersage“ zufälliger Koeffizienten in gemischten linearen Modellen erweiterte. Die folgende Darstellung des BLUP's für  $\beta_i$  im RCR-Modell stammt aus einer Veröffentlichung von Bondeson (1990). Die Beweis-idee ist Frees (2006, S. 157 ff.) entnommen.

**Satz 3.27: Geschlossene Darstellung des BLUP's für  $\lambda'\beta_i$  im RCR-Modell**

Sei  $Y_i = X\beta_i + \epsilon_i$ ,  $i \in \{1, \dots, I\}$ , das in Definition 3.17 vorgestellte RCR-Modell, für welches die Annahmen [R1]-[R4] gelten. Sei  $W := \lambda'\beta_i = \lambda'b + \lambda'\delta_i$ ,  $\lambda \in \mathbb{R}^k$ , eine Linearkombination aus festen und zufälligen Effekten. Dann ist der beste lineare unverzerrte Prädiktor (BLUP) für  $W$  gegeben durch

$$\widehat{W}_{\text{BLUP}} := \lambda' \left( C\widehat{\beta}_{i,\text{OLS}} + (I_k - C)\widehat{b}_{\text{GLS}} \right).$$

Hierbei ist  $I_k$  die  $(k \times k)$ -Einheitsmatrix. Die Matrix  $C$  ist definiert als  $C := \Sigma (\Sigma + \sigma^2(X'X)^{-1})^{-1}$ .

**Beweis:**

$\widehat{W}_{\text{BLUP}}$  ist ein linearer Prädiktor, da sowohl  $\widehat{\beta}_{i,\text{OLS}}$  als auch  $\widehat{b}_{\text{GLS}}$  linear in  $Y$  sind. Zudem ist  $\widehat{W}_{\text{BLUP}}$  unverzerrt, da gilt:

$$\begin{aligned} \mathbb{E} \left( \widehat{W}_{\text{BLUP}} \right) &= \mathbb{E} \left( \lambda' \left( C\widehat{\beta}_{i,\text{OLS}} + (I_k - C)\widehat{b}_{\text{GLS}} \right) \right) \\ &= \lambda' C \mathbb{E} \left( \widehat{\beta}_{i,\text{OLS}} \right) + \lambda' (I_k - C) \mathbb{E} \left( \widehat{b}_{\text{GLS}} \right) \end{aligned}$$

$$\begin{aligned}
&= \lambda' \mathbf{C} \mathbf{b} + \lambda' (\mathbf{I}_k - \mathbf{C}) \mathbf{b} \quad (\text{nach Satz 3.19 und 3.21}) \\
&= \lambda' (\mathbf{C} + \mathbf{I}_k - \mathbf{C}) \mathbf{b} \\
&= \lambda' \mathbf{b} \\
&= \mathbb{E}(W)
\end{aligned}$$

Nun wird gezeigt, dass  $\widehat{W}_{\text{BLUP}}$  den kleinsten mittleren quadratischen Fehler innerhalb der Klasse der linearen, unverzerrten Prädiktoren besitzt. Dazu sei

$$\mathcal{W} := \{ \widehat{W} := \widehat{W}_{\text{BLUP}} + \mathbf{a}' \mathbf{Y} \mid \mathbf{a} \in \mathbb{R}^{NI} \text{ mit } \mathbb{E}(\mathbf{a}' \mathbf{Y}) = 0 \}$$

die Klasse der linearen unverzerrten Prädiktoren für  $W$ . Die Bedingung  $\mathbb{E}(\mathbf{a}' \mathbf{Y}) = 0$  muss erfüllt sein, damit die Elemente in  $\mathcal{W}$  erwartungstreu für  $W$  sind. Diese Bedingung ist für alle  $\mathbf{a} \in \mathbb{R}^{NI}$  erfüllt, die der Gleichung  $\mathbf{a}' \mathcal{X} = \mathbf{0}$  genügen. Mit den im Beweis zu Satz 3.21 eingeführten Notationen lässt sich dieses Argument durch folgende Umformung begründen:

$$\begin{aligned}
&\mathbb{E}(\mathbf{a}' \mathbf{Y}) = 0 \\
\Leftrightarrow &\mathbb{E}(\mathbf{a}' (\mathcal{X} \mathbf{b} + \mathbf{D}(\mathbf{X}) \boldsymbol{\delta} + \boldsymbol{\epsilon})) = 0 \\
\Leftrightarrow &\mathbf{a}' \mathcal{X} \mathbf{b} + \mathbf{a}' \mathbf{D}(\mathbf{X}) \mathbb{E}(\boldsymbol{\delta}) + \mathbf{a}' \mathbb{E}(\boldsymbol{\epsilon}) = 0 \\
\Leftrightarrow &\mathbf{a}' \mathcal{X} \mathbf{b} = 0 \quad (\text{nach [R2] und [R3]})
\end{aligned}$$

Da  $\mathbf{b} \neq \mathbf{0}$  angenommen wird, muss also  $\mathbf{a}' \mathcal{X} = \mathbf{0}$  sein. Nun soll  $\mathbf{a} \in \mathbb{R}^{NI}$  so gewählt werden, dass der mittlere quadratische Fehler  $\mathbb{E} \left( \left( \widehat{W}_{\text{BLUP}} + \mathbf{a}' \mathbf{Y} - W \right)^2 \right)$  minimiert wird. Dazu wird der mittlere quadratische Fehler zunächst in Komponenten zerlegt:

$$\begin{aligned}
\mathbb{E} \left( \left( \widehat{W}_{\text{BLUP}} + \mathbf{a}' \mathbf{Y} - W \right)^2 \right) &= \mathbb{E} \left( \left( \widehat{W}_{\text{BLUP}} - W + \mathbf{a}' \mathbf{Y} \right)^2 \right) \\
&= \mathbb{V}\text{ar} \left( \widehat{W}_{\text{BLUP}} - W + \mathbf{a}' \mathbf{Y} \right) + \mathbb{E} \left( \widehat{W}_{\text{BLUP}} - W + \mathbf{a}' \mathbf{Y} \right)^2 \\
&= \mathbb{V}\text{ar} \left( \widehat{W}_{\text{BLUP}} - W + \mathbf{a}' \mathbf{Y} \right) + (\lambda' \mathbf{b} - \lambda' \mathbf{b} + 0)^2 \\
&= \mathbb{C}\text{ov} \left( \widehat{W}_{\text{BLUP}} - W + \mathbf{a}' \mathbf{Y}, \widehat{W}_{\text{BLUP}} - W + \mathbf{a}' \mathbf{Y} \right) \\
&= \mathbb{V}\text{ar} \left( \widehat{W}_{\text{BLUP}} - W \right) + \mathbb{V}\text{ar} \left( \mathbf{a}' \mathbf{Y} \right) + 2 \mathbb{C}\text{ov} \left( \widehat{W}_{\text{BLUP}} - W, \mathbf{a}' \mathbf{Y} \right)
\end{aligned}$$

Der Kovarianzterm in der letzten Gleichung fällt weg, da  $\mathbb{C}\text{ov} \left( \widehat{W}_{\text{BLUP}} - W, \mathbf{a}' \mathbf{Y} \right) = 0$  gilt. Dies wird durch die folgenden Umformungen gezeigt. Wie im Beweis zu Satz 3.21 gilt bezüglich der Notationen wieder  $\mathcal{X} := (\mathbf{X}', \dots, \mathbf{X}')' \in \mathbb{R}^{NI \times k}$ ,  $\mathbf{Y} := (\mathbf{Y}_1', \dots, \mathbf{Y}_I')' \in \mathbb{R}^{NI}$  und  $\mathbf{H} = \mathbb{V}\text{ar}(\mathbf{Y}) \in \mathbb{R}^{NI \times NI}$ . Desweiteren sei  $\mathbf{e}_I^{(i)}$  der  $i$ -te Einheitsvektor des  $\mathbb{R}^I$ :

$$\begin{aligned}
& \text{Cov} \left( \widehat{W}_{\text{BLUP}} - W, \mathbf{a}'\mathbf{Y} \right) \\
&= \text{Cov} \left( \widehat{W}_{\text{BLUP}}, \mathbf{Y} \right) \mathbf{a} - \text{Cov} (W, \mathbf{Y}) \mathbf{a} \\
&= \text{Cov} \left( \lambda' \mathbf{C} \widehat{\boldsymbol{\beta}}_{i,\text{OLS}} + \lambda' (\mathbf{I}_k - \mathbf{C}) \widehat{\mathbf{b}}_{\text{GLS}}, \mathbf{Y} \right) \mathbf{a} - \text{Cov} (\lambda' \mathbf{b} + \lambda' \boldsymbol{\delta}_i, \mathbf{X} \mathbf{b} + \mathbf{D}(\mathbf{X}) \boldsymbol{\delta} + \boldsymbol{\epsilon}) \mathbf{a} \\
&= \text{Cov} \left( \lambda' \mathbf{C} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \mathbf{Y}_i, \mathbf{Y} \right) \mathbf{a} + \text{Cov} \left( \lambda' (\mathbf{I}_k - \mathbf{C}) \left( \mathbf{X}' \mathbf{H}^{-1} \mathbf{X} \right)^{-1} \mathbf{X}' \mathbf{H}^{-1} \mathbf{Y}, \mathbf{Y} \right) \mathbf{a} \\
&\quad - \text{Cov} (\lambda' \boldsymbol{\delta}_i, \mathbf{D}(\mathbf{X}) \boldsymbol{\delta}) \mathbf{a} - \text{Cov} (\lambda' \boldsymbol{\delta}_i, \boldsymbol{\epsilon}) \mathbf{a} \\
&= \lambda' \mathbf{C} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \text{Cov} (\mathbf{Y}_i, \mathbf{Y}) \mathbf{a} + \lambda' (\mathbf{I}_k - \mathbf{C}) \left( \mathbf{X}' \mathbf{H}^{-1} \mathbf{X} \right)^{-1} \mathbf{X}' \mathbf{H}^{-1} \text{Var}(\mathbf{Y}) \mathbf{a} \\
&\quad - \lambda' \text{Cov}(\boldsymbol{\delta}_i, \boldsymbol{\delta}) \mathbf{D}(\mathbf{X})' \mathbf{a} \quad (\text{nach [R4]}) \\
&= \lambda' \boldsymbol{\Sigma} \left( \boldsymbol{\Sigma} + \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} \right)^{-1} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \left( \mathbf{e}_I^{(i)'} \otimes \left( \mathbf{X} \boldsymbol{\Sigma} \mathbf{X}' + \sigma^2 \mathbf{I}_N \right) \right) \mathbf{a} \\
&\quad + \lambda' (\mathbf{I}_k - \mathbf{C}) \left( \mathbf{X}' \mathbf{H}^{-1} \mathbf{X} \right)^{-1} \mathbf{X}' \mathbf{H}^{-1} \mathbf{H} \mathbf{a} - \lambda' \left( \mathbf{e}_I^{(i)'} \otimes \boldsymbol{\Sigma} \right) \mathbf{D}(\mathbf{X})' \mathbf{a} \quad (\text{Satz 3.18 und [R2]}) \\
&= \lambda' \left[ \mathbf{e}_I^{(i)'} \otimes \left( \boldsymbol{\Sigma} \left( \boldsymbol{\Sigma} + \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} \right)^{-1} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \left( \mathbf{X} \boldsymbol{\Sigma} \mathbf{X}' + \sigma^2 \mathbf{I}_N \right) \right) \right] \mathbf{a} \\
&\quad + \lambda' (\mathbf{I}_k - \mathbf{C}) \left( \mathbf{X}' \mathbf{H}^{-1} \mathbf{X} \right)^{-1} \mathbf{X}' \mathbf{H}^{-1} \mathbf{H} \mathbf{a} - \lambda' \left( \mathbf{e}_I^{(i)'} \otimes \boldsymbol{\Sigma} \right) \mathbf{D}(\mathbf{X})' \mathbf{a} \\
&= \lambda' \left[ \mathbf{e}_I^{(i)'} \otimes \left( \boldsymbol{\Sigma} \left( \boldsymbol{\Sigma} + \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} \right)^{-1} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \mathbf{X} \boldsymbol{\Sigma} \mathbf{X}' + \sigma^2 \boldsymbol{\Sigma} \left( \boldsymbol{\Sigma} + \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} \right)^{-1} \right. \right. \\
&\quad \left. \left. \times (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \right) \right] \mathbf{a} + \lambda' (\mathbf{I}_k - \mathbf{C}) \left( \mathbf{X}' \mathbf{H}^{-1} \mathbf{X} \right)^{-1} \mathbf{X}' \mathbf{H}^{-1} \mathbf{H} \mathbf{a} - \lambda' \left( \mathbf{e}_I^{(i)'} \otimes \boldsymbol{\Sigma} \right) \mathbf{D}(\mathbf{X})' \mathbf{a} \\
&= \lambda' \left[ \mathbf{e}_I^{(i)'} \otimes \left( \boldsymbol{\Sigma} \left( \boldsymbol{\Sigma} + \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} \right)^{-1} \left( \boldsymbol{\Sigma} + \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} \right) \mathbf{X}' \right) \right] \mathbf{a} \\
&\quad + \lambda' (\mathbf{I}_k - \mathbf{C}) \left( \mathbf{X}' \mathbf{H}^{-1} \mathbf{X} \right)^{-1} \mathbf{X}' \mathbf{H}^{-1} \mathbf{H} \mathbf{a} - \lambda' \left( \mathbf{e}_I^{(i)'} \otimes \boldsymbol{\Sigma} \right) \mathbf{D}(\mathbf{X})' \mathbf{a} \\
&= \lambda' \left[ \mathbf{e}_I^{(i)'} \otimes (\boldsymbol{\Sigma} \mathbf{X}') \right] \mathbf{a} + \lambda' (\mathbf{I}_k - \mathbf{C}) \left( \mathbf{X}' \mathbf{H}^{-1} \mathbf{X} \right)^{-1} \mathbf{X}' \mathbf{H}^{-1} \mathbf{H} \mathbf{a} \\
&\quad - \lambda' \left( \mathbf{e}_I^{(i)'} \otimes \boldsymbol{\Sigma} \right) \mathbf{D}(\mathbf{X})' \mathbf{a} \\
&= \lambda' \left( \mathbf{e}_I^{(i)'} \otimes (\boldsymbol{\Sigma} \mathbf{X}') \right) \mathbf{a} + \lambda' (\mathbf{I}_k - \mathbf{C}) \left( \mathbf{X}' \mathbf{H}^{-1} \mathbf{X} \right)^{-1} (\mathbf{a}' \mathbf{X})' \\
&\quad - \lambda' \left[ \mathbf{e}_I^{(i)'} \otimes (\boldsymbol{\Sigma} \mathbf{X}') \right] \mathbf{a} \\
&= 0, \text{ da } \mathbf{a}' \mathbf{X} = \mathbf{0} \text{ nach Voraussetzung}
\end{aligned}$$



Somit kann der mittlere quadratische Fehler eines beliebigen linearen unverzerrten Prädiktors  $\widehat{W}$  für  $W$  ausgedrückt werden als

$$\text{MSE}(\widehat{W}, W) = \mathbb{E} \left( \left( \widehat{W}_{\text{BLUP}} + \mathbf{a}'\mathbf{Y} - W \right)^2 \right) = \text{Var}(\widehat{W}_{\text{BLUP}} - W) + \text{Var}(\mathbf{a}'\mathbf{Y}) .$$

Da  $\text{Var}(\mathbf{a}'\mathbf{Y}) > 0 \forall \mathbf{a} \neq \mathbf{0}$ , wird der mittlere quadratische Fehler durch die Wahl von  $\mathbf{a} = \mathbf{0}$  minimiert. Somit ist  $\widehat{W}_{\text{BLUP}} := \lambda' \left( \mathbf{C}\widehat{\boldsymbol{\beta}}_{i,\text{OLS}} + (\mathbf{I}_k - \mathbf{C})\widehat{\mathbf{b}}_{\text{GLS}} \right)$  der beste lineare unverzerrte Prädiktor für  $W$ .  $\square$

Im vektorwertigen Sinn ist der BLUP für  $\boldsymbol{\beta}_i = \mathbf{b} + \boldsymbol{\delta}_i$  somit gegeben durch eine gewichtete Summe aus dem OLS-Schätzer  $\widehat{\boldsymbol{\beta}}_{i,\text{OLS}}$  und dem GLS-Schätzer  $\widehat{\mathbf{b}}_{\text{GLS}}$ :

$$\widehat{\boldsymbol{\beta}}_{i,\text{BLUP}} := \mathbf{C}\widehat{\boldsymbol{\beta}}_{i,\text{OLS}} + (\mathbf{I}_k - \mathbf{C})\widehat{\mathbf{b}}_{\text{GLS}} , \text{ wobei } \mathbf{C} := \boldsymbol{\Sigma} \left( \boldsymbol{\Sigma} + \sigma^2(\mathbf{X}'\mathbf{X})^{-1} \right)^{-1}$$

Intuitiv würde man vermuten, dass  $\widehat{\boldsymbol{\beta}}_{i,\text{OLS}}$  der optimale Prädiktor für  $\boldsymbol{\beta}_i$  ist. Dies ist jedoch nicht der Fall: Für eine optimale Prädiktion des Koeffizientenvektors  $\boldsymbol{\beta}_i$  wird die Information aller Verläufe  $\mathbf{Y}_1, \dots, \mathbf{Y}_I$  benötigt und nicht nur die Information des  $i$ -ten Verlaufs  $\mathbf{Y}_i$ . Das Gewicht liegt dabei umso stärker auf dem OLS-Schätzer  $\widehat{\boldsymbol{\beta}}_{i,\text{OLS}}$ , je größer die Varianzen der zufälligen Effekte sind und je kleiner die Designvarianz  $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$  ist. In diesem Fall gilt nämlich approximativ:

$$\mathbf{C} = \boldsymbol{\Sigma} \left( \boldsymbol{\Sigma} + \sigma^2(\mathbf{X}'\mathbf{X})^{-1} \right)^{-1} \approx \boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1} = \mathbf{I}_k$$

Dies bedeutet: Je größer die interindividuelle Varianz zwischen den Verläufen ist und je mehr Beobachtungen pro Verlauf vorliegen, desto stärker liegt das Gewicht auf dem OLS-Schätzer  $\widehat{\boldsymbol{\beta}}_{i,\text{OLS}}$  (Kadiyala und Oberhelman, 1982).

Das Problem bei der Anwendung von BLUP's besteht darin, dass sie von den Varianzkomponenten  $\sigma^2$  und  $\boldsymbol{\Sigma}$  abhängen. Diese sind in der Praxis nicht bekannt. Daher werden  $\sigma^2$  und  $\boldsymbol{\Sigma}$  durch ihre Schätzer  $\widehat{\sigma}_{\text{RCR}}^2$  und  $\widehat{\boldsymbol{\Sigma}}$  ersetzt. Falls in der Formel eines BLUP's die geschätzten anstatt der wahren Varianzkomponenten eingesetzt werden, bezeichnet man diese Prädiktoren als „empirische beste lineare unverzerrte Prädiktoren“ (EBLUP's). Durch die Schätzung der Varianzkomponenten verliert der Prädiktor allerdings seine Optimalitätseigenschaften (Frees, 2006, S. 135 ff.). Der EBLUP für  $\boldsymbol{\beta}_i$  ist somit gegeben durch

$$\widehat{\boldsymbol{\beta}}_{i,\text{EBLUP}} := \widehat{\mathbf{C}}\widehat{\boldsymbol{\beta}}_{i,\text{OLS}} + \left( \mathbf{I}_k - \widehat{\mathbf{C}} \right) \widehat{\mathbf{b}}_{\text{GLS}} , \text{ wobei } \widehat{\mathbf{C}} := \widehat{\boldsymbol{\Sigma}} \left( \widehat{\boldsymbol{\Sigma}} + \widehat{\sigma}_{\text{RCR}}^2(\mathbf{X}'\mathbf{X})^{-1} \right)^{-1} .$$

Mithilfe der empirischen BLUP's können nun auch die verlaufsspezifischen Störgrößen  $\boldsymbol{\epsilon}_i$ ,  $i \in \{1, \dots, I\}$ , durch  $\widehat{\boldsymbol{\epsilon}}_{i,\text{EBLUP}} := \mathbf{Y}_i - \mathbf{X}\widehat{\boldsymbol{\beta}}_{i,\text{EBLUP}}$  vorhergesagt werden. Diese werden als BLUP-Residuen bezeichnet (Frees, 2006, S. 135 ff.).

### 3.4.3 Prognosen und Prognoseintervalle für zukünftige Beobachtungen im RCR-Modell

Im Kontext des RCR-Modells für longitudinale Daten kommen verschiedene Vorhersageszenarien in Betracht, für die in diesem Abschnitt Prognosen und Prognoseintervalle hergeleitet werden. Sei  $T$  im Folgenden der Zeitpunkt, für den eine neue Beobachtung vorhergesagt werden soll und  $\mathbf{x}_T := (1, T, T^2, \dots)'$  der entsprechende Vektor der unabhängigen Variablen.

#### 3.4.3.1 Situation A: Prognosen für einen neuen, unbekanntem Verlauf

Es liegen balancierte, longitudinale Daten von  $I$  Verläufen vor, für welche die gesamten Messzeitpunkte gleich sind. Anhand dieser Daten wird das in Definition 3.17 vorgestellte RCR-Modell angepasst (d.h.  $\mathbf{b}$ ,  $\sigma^2$  und  $\Sigma$  werden geschätzt). Sei  $Y_{0T} = \mathbf{x}_T' \boldsymbol{\beta}_0 + \epsilon_{0T}$  die unbekannte Beobachtung eines neuen,  $(I + 1)$ -ten Verlaufs zum Zeitpunkt  $T$ . Von diesem neuen Verlauf (mit Index 0) liegen noch keine Beobachtungen vor. Der neue Verlauf folge den für das RCR-Modell gemachten Annahmen: Es gelte also  $\boldsymbol{\beta}_0 \sim \mathcal{N}(\mathbf{b}, \Sigma)$ ,  $\epsilon_{0T} \sim \mathcal{N}(0, \sigma^2)$  und die stochastische Unabhängigkeit zwischen  $\epsilon_{0T}, \epsilon_1, \dots, \epsilon_I, \boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_I$ . Dies bedeutet insbesondere, dass die Beobachtung  $Y_{0T}$  stochastisch unabhängig vom Zufallsvektor  $\mathbf{Y} = (\mathbf{Y}_1', \dots, \mathbf{Y}_I')'$  ist. In dieser Situation ist die beste lineare unverzerrte Vorhersage für  $Y_{0T}$  nach Swamy (1971, S. 133 ff.) gegeben durch

$$\hat{Y}_{0T} := \mathbf{x}_T' \hat{\mathbf{b}}_{\text{GLS}},$$

da  $\hat{\mathbf{b}}_{\text{GLS}}$  der beste lineare unverzerrte Schätzer der festen Effekte ist und da über den neuen Verlauf keinerlei Informationen vorliegen. Im Folgenden wird ein Prognoseintervall für  $Y_{0T}$  hergeleitet. Zunächst wird dazu die Verteilung des Prognosefehlers betrachtet.

#### Lemma 3.28: Verteilung des Prognosefehlers für Situation A

Sei  $\mathbf{Y}_i = \mathbf{X}\boldsymbol{\beta}_i + \boldsymbol{\epsilon}_i$ ,  $i \in \{1, \dots, I\}$ , das in Definition 3.17 vorgestellte RCR-Modell, für welches die Annahmen [R1]-[R5] erfüllt sind. Sei  $\hat{Y}_{0T} := \mathbf{x}_T' \hat{\mathbf{b}}_{\text{GLS}}$  die Prognose für die Beobachtung eines neuen, bislang völlig unbekanntem Verlaufs zum Zeitpunkt  $T$ . Dann ist der Prognosefehler  $(\hat{Y}_{0T} - Y_{0T})$  normalverteilt mit Erwartungswert 0 und Varianz

$$v := \frac{1}{I} \mathbf{x}_T' \left( \Sigma + \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} \right) \mathbf{x}_T + \mathbf{x}_T' \Sigma \mathbf{x}_T + \sigma^2.$$

Es gilt also  $(\hat{Y}_{0T} - Y_{0T}) \sim \mathcal{N}(0, v)$ .

**Beweis:**

Da  $\hat{Y}_{0T}$  und  $Y_{0T}$  normalverteilt sind, ist auch  $(\hat{Y}_{0T} - Y_{0T})$  normalverteilt. Für den Erwartungswert und die Varianz des Prognosefehlers gilt:

$$\begin{aligned}\mathbb{E}(\hat{Y}_{0T} - Y_{0T}) &= \mathbb{E}(\hat{Y}_{0T}) - \mathbb{E}(Y_{0T}) \\ &= \mathbb{E}(\mathbf{x}'_T \hat{\mathbf{b}}_{\text{GLS}}) - \mathbb{E}(\mathbf{x}'_T \boldsymbol{\beta}_0 + \epsilon_{0T}) \\ &= \mathbf{x}'_T \mathbb{E}(\hat{\mathbf{b}}_{\text{GLS}}) - \mathbf{x}'_T \mathbb{E}(\boldsymbol{\beta}_0) - \mathbb{E}(\epsilon_{0T}) \\ &= \mathbf{x}'_T \mathbf{b} - \mathbf{x}'_T \mathbf{b} \quad (\text{nach Satz 3.21 und [R2], [R3]}) \\ &= 0\end{aligned}$$

$$\begin{aligned}\text{Var}(\hat{Y}_{0T} - Y_{0T}) &= \text{Var}(\hat{Y}_{0T}) + \text{Var}(Y_{0T}) \quad (\text{da } Y_{0T} \text{ unabhg. von } \mathbf{Y}) \\ &= \text{Var}(\mathbf{x}'_T \hat{\mathbf{b}}_{\text{GLS}}) + \text{Var}(\mathbf{x}'_T \boldsymbol{\beta}_0 + \epsilon_{0T}) \\ &= \mathbf{x}'_T \text{Var}(\hat{\mathbf{b}}_{\text{GLS}}) \mathbf{x}_T + \mathbf{x}'_T \text{Var}(\boldsymbol{\beta}_0) \mathbf{x}_T + \text{Var}(\epsilon_{0T}) \quad (\text{nach [R4]}) \\ &= \frac{1}{I} \mathbf{x}'_T (\boldsymbol{\Sigma} + \sigma^2 (\mathbf{X}'\mathbf{X})^{-1}) \mathbf{x}_T + \mathbf{x}'_T \boldsymbol{\Sigma} \mathbf{x}_T + \sigma^2 \quad (\text{nach Satz 3.21 und [R2], [R3]})\end{aligned}$$

□

Falls die Varianzkomponenten  $\boldsymbol{\Sigma}$  und  $\sigma^2$  bekannt wären, könnte bereits aus dem standardisierten Prognosefehler ein Prognoseintervall hergeleitet werden, da gilt:

$$\frac{\hat{Y}_{0T} - Y_{0T}}{\sqrt{v}} \sim \mathcal{N}(0, 1).$$

Die Prognosevarianz  $v$  ist jedoch nicht bekannt, da  $\boldsymbol{\Sigma}$  und  $\sigma^2$  unbekannt sind. Aus diesem Grund muss  $v$  geschätzt werden, indem die Schätzer  $\hat{\boldsymbol{\Sigma}}$  und  $\hat{\sigma}^2$  in die Formel eingesetzt werden. Mithilfe dieses Prognosevarianz-Schätzers kann ein approximatives Prognoseintervall für  $Y_{0T}$  hergeleitet werden. Dazu wird ein Hilfssatz benötigt, der eine allgemeine Aussage zur Konvergenz von Zufallsvariablen macht.

**Lemma 3.29: Theorem von Slutsky**

Sei  $\{Z_n, n \in \mathbb{N}\}$  eine Folge von Zufallsvariablen, die in Verteilung gegen die Zufallsvariable  $Z$  konvergiert. Seien  $\{A_n, n \in \mathbb{N}\}$  und  $\{B_n, n \in \mathbb{N}\}$  Folgen von Zufallsvariablen, die in Wahrscheinlichkeit gegen die Konstanten  $a$  bzw.  $b$  konvergieren. Dann konvergiert die Funktion  $(A_n + B_n Z_n)$  für  $n \rightarrow \infty$  in Verteilung gegen  $(a + bZ)$  (Rao, 1973, S. 122). Kurz:

$$Z_n \xrightarrow{D} Z, A_n \xrightarrow{P} a, B_n \xrightarrow{P} b \implies A_n + B_n Z_n \xrightarrow{D} a + bZ$$

**Satz 3.30: Approximatives Prognoseintervall für  $Y_{0T}$  in Situation A**

Sei  $Y_i = \mathbf{X}\beta_i + \varepsilon_i$ ,  $i \in \{1, \dots, I\}$ , das in Definition 3.17 vorgestellte RCR-Modell, für welches die Annahmen [R1]-[R5] erfüllt sind. Sei  $\hat{Y}_{0T} := \mathbf{x}'_T \hat{\mathbf{b}}_{\text{GLS}}$  die Prognose der Beobachtung eines neuen, völlig unbekanntem Verlaufs zum Zeitpunkt  $T$ . Dann kann die Varianz  $v$  des Prognosefehlers durch

$$\hat{v} := \frac{1}{I} \mathbf{x}'_T \left( \hat{\Sigma} + \hat{\sigma}_{\text{RCR}}^2 (\mathbf{X}'\mathbf{X})^{-1} \right) \mathbf{x}_T + \mathbf{x}'_T \hat{\Sigma} \mathbf{x}_T + \hat{\sigma}_{\text{RCR}}^2$$

erwartungstreu geschätzt werden und es gilt: Der durch die geschätzte Prognosevarianz standardisierte Prognosefehler konvergiert für  $N, I \rightarrow \infty$  in Verteilung gegen die Standardnormalverteilung:

$$\frac{\hat{Y}_{0T} - Y_{0T}}{\sqrt{\hat{v}}} \xrightarrow[N \rightarrow \infty, I \rightarrow \infty]{D} \mathcal{N}(0, 1)$$

Daraus ergibt sich folgendes  $(1 - \alpha)$ -Prognoseintervall für die Beobachtung  $Y_{0T}$ :

$$\widehat{\mathcal{PI}}(\mathbf{Y}) = \left[ \hat{Y}_{0T} - q_{1-\alpha/2} \sqrt{\hat{v}}, \hat{Y}_{0T} + q_{1-\alpha/2} \sqrt{\hat{v}} \right]$$

Hierbei ist  $q_{1-\alpha/2}$  das  $(1 - \alpha/2)$ -Quantil der Standardnormalverteilung.

**Beweis:**

Da  $\hat{\Sigma}$  und  $\hat{\sigma}_{\text{RCR}}^2$  erwartungstreue Schätzer für  $\Sigma$  bzw.  $\sigma^2$  sind, ist auch  $\hat{v}$  ein erwartungstreuer Schätzer der Prognosevarianz  $v$ . Um die asymptotischen Eigenschaften des Schätzers  $\hat{v}$  zu betrachten, wird zuerst  $N$  und dann  $I$  gegen  $\infty$  streben gelassen:

$$\begin{aligned} \hat{v}^* &:= \text{plim}_{N \rightarrow \infty}(\hat{v}) = \text{plim}_{N \rightarrow \infty} \left( \frac{1}{I} \mathbf{x}'_T \left( \hat{\Sigma} + \hat{\sigma}_{\text{RCR}}^2 (\mathbf{X}'\mathbf{X})^{-1} \right) \mathbf{x}_T + \mathbf{x}'_T \hat{\Sigma} \mathbf{x}_T + \hat{\sigma}_{\text{RCR}}^2 \right) \\ &= \frac{1}{I} \mathbf{x}'_T \text{plim}_{N \rightarrow \infty} \left( \hat{\Sigma} \right) \mathbf{x}_T + \frac{1}{I} \mathbf{x}'_T \text{plim}_{N \rightarrow \infty} \left( \hat{\sigma}_{\text{RCR}}^2 (\mathbf{X}'\mathbf{X})^{-1} \right) \mathbf{x}_T + \mathbf{x}'_T \text{plim}_{N \rightarrow \infty} \left( \hat{\Sigma} \right) \mathbf{x}_T \\ &\quad + \text{plim}_{N \rightarrow \infty} \left( \hat{\sigma}_{\text{RCR}}^2 \right) \\ &= \frac{1}{I} \mathbf{x}'_T \left( \frac{1}{I-1} \mathbf{S} \right) \mathbf{x}_T + \mathbf{x}'_T \left( \frac{1}{I-1} \mathbf{S} \right) \mathbf{x}_T + \sigma^2 \quad (\text{nach Satz 3.22 und 3.23}) \\ &= \frac{I+1}{I} \mathbf{x}'_T \left( \frac{1}{I-1} \mathbf{S} \right) \mathbf{x}_T + \sigma^2 \end{aligned}$$

$$\begin{aligned} \text{plim}_{I \rightarrow \infty}(\hat{v}^*) &= \text{plim}_{I \rightarrow \infty} \left( \frac{I+1}{I} \mathbf{x}'_T \left( \frac{1}{I-1} \mathbf{S} \right) \mathbf{x}_T + \sigma^2 \right) \\ &= \lim_{I \rightarrow \infty} \left( \frac{I+1}{I} \right) \mathbf{x}'_T \text{plim}_{I \rightarrow \infty} \left( \frac{1}{I-1} \mathbf{S} \right) \mathbf{x}_T + \sigma^2 \\ &= \mathbf{x}'_T \Sigma \mathbf{x}_T + \sigma^2 \quad (\text{nach Satz 3.23}) \end{aligned}$$

Der Prognosevarianzschätzer  $\hat{v}$  ist also ein konsistenter Schätzer für  $(\mathbf{x}'_T \boldsymbol{\Sigma} \mathbf{x}_T + \sigma^2)$ , nicht jedoch für die exakte Prognosevarianz  $v$ . Daher stellt sich die Frage, gegen welche Verteilung der Prognosefehler  $(\hat{Y}_{0T} - Y_{0T})$  für  $N, I \rightarrow \infty$  konvergiert. Da der Erwartungswert des Prognosefehlers 0 ist – also nicht von  $I$  und  $N$  abhängt – wird nun betrachtet, gegen welchen Wert die wahre Prognosevarianz  $v$  konvergiert. Es gelte wieder, dass der Grenzwert  $\mathbf{M} := \lim_{N \rightarrow \infty} (N(\mathbf{X}'\mathbf{X})^{-1})$  existiert und positiv definit ist.

$$\begin{aligned} v^* &:= \lim_{N \rightarrow \infty} v = \lim_{N \rightarrow \infty} \left( \frac{1}{I} \mathbf{x}'_T \left( \boldsymbol{\Sigma} + \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} \right) \mathbf{x}_T + \mathbf{x}'_T \boldsymbol{\Sigma} \mathbf{x}_T + \sigma^2 \right) \\ &= \frac{I+1}{I} \mathbf{x}'_T \boldsymbol{\Sigma} \mathbf{x}_T + \left( \frac{\sigma^2}{I} \right) \mathbf{x}'_T \lim_{N \rightarrow \infty} \left( \frac{1}{N} \right) \lim_{N \rightarrow \infty} \left( N(\mathbf{X}'\mathbf{X})^{-1} \right) \mathbf{x}_T + \sigma^2 \\ &= \frac{I+1}{I} \mathbf{x}'_T \boldsymbol{\Sigma} \mathbf{x}_T + \left( \frac{\sigma^2}{I} \right) \mathbf{x}'_T (0 \cdot \mathbf{M}) \mathbf{x}_T + \sigma^2 \\ &= \frac{I+1}{I} \mathbf{x}'_T \boldsymbol{\Sigma} \mathbf{x}_T + \sigma^2 \end{aligned}$$

$$\begin{aligned} \lim_{I \rightarrow \infty} v^* &= \lim_{I \rightarrow \infty} \left( \frac{I+1}{I} \right) \mathbf{x}'_T \boldsymbol{\Sigma} \mathbf{x}_T + \sigma^2 \\ &= \mathbf{x}'_T \boldsymbol{\Sigma} \mathbf{x}_T + \sigma^2 \end{aligned}$$

Daraus folgt, dass der Prognosefehler  $(\hat{Y}_{0T} - Y_{0T})$  für  $N, I \rightarrow \infty$  in Verteilung gegen eine Normalverteilung mit Erwartungswert 0 und Varianz  $(\mathbf{x}'_T \boldsymbol{\Sigma} \mathbf{x}_T + \sigma^2)$  konvergiert:

$$(\hat{Y}_{0T} - Y_{0T}) \xrightarrow[N \rightarrow \infty, I \rightarrow \infty]{D} Z \sim \mathcal{N}(0, \mathbf{x}'_T \boldsymbol{\Sigma} \mathbf{x}_T + \sigma^2)$$

Da der Prognosevarianzschätzer  $\hat{v}$  in Wahrscheinlichkeit gegen die Varianz der Grenzverteilung  $Z$  konvergiert, folgt mit dem Theorem von Slutsky (Lemma 3.29):

$$\frac{\hat{Y}_{0T} - Y_{0T}}{\sqrt{\hat{v}}} \xrightarrow[N \rightarrow \infty, I \rightarrow \infty]{D} \frac{Z}{\sqrt{\text{Var}(Z)}} \sim \mathcal{N}(0, 1)$$

Der mit der geschätzten Prognosevarianz standardisierte Prognosefehler konvergiert also in Verteilung gegen die Standardnormalverteilung. Ein approximatives  $(1 - \alpha)$ -Prognoseintervall für die Beobachtung  $Y_{0T}$  eines neuen, völlig unbekanntem Verlaufs ist somit gegeben durch

$$\left[ \hat{Y}_{0T} - q_{1-\alpha/2} \sqrt{\hat{v}}, \hat{Y}_{0T} + q_{1-\alpha/2} \sqrt{\hat{v}} \right].$$

□

### 3.4.3.2 Situation B: Prognosen für einen vollständig beobachteten Verlauf

Es liegen balancierte, longitudinale Daten von  $I$  Verläufen vor, für welche die gesamten Messzeitpunkte gleich sind. Anhand dieser Daten wird das in Definition 3.17 vorgestellte RCR-Modell angepasst. Nun besteht das Interesse darin, für einen der  $I$  Verläufe eine Beobachtung zu einem zukünftigen Zeitpunkt  $T$ ,  $T > t_N$ , zu prognostizieren. Der Index dieses Verlaufs sei  $i$ ,  $i \in \{1, \dots, I\}$ . Der empirische beste lineare unverzerrte Prädiktor für die Beobachtung  $Y_{iT}$  ist dann nach Bondeson (1990) gegeben durch

$$\hat{Y}_{iT, \text{EBLUP}} = \mathbf{x}'_T \hat{\boldsymbol{\beta}}_{i, \text{EBLUP}} = \mathbf{x}'_T \left( \hat{\mathbf{C}} \hat{\boldsymbol{\beta}}_{i, \text{OLS}} + (\mathbf{I}_k - \hat{\mathbf{C}}) \hat{\mathbf{b}}_{\text{GLS}} \right),$$

wobei  $\hat{\mathbf{C}} := \hat{\boldsymbol{\Sigma}} \left( \hat{\boldsymbol{\Sigma}} + \hat{\sigma}_{\text{RCR}}^2 (\mathbf{X}'\mathbf{X})^{-1} \right)^{-1}$  ist.

### 3.4.3.3 Situation C: Prognosen für einen neuen Verlauf, von dem nur die ersten Messungen bekannt sind

Es liegen balancierte, longitudinale Daten von  $I$  Verläufen vor, für welche die gesamten Messzeitpunkte gleich sind. Anhand dieser Daten wird das in Definition 3.17 vorgestellte RCR-Modell angepasst. Von einem neuen,  $(I + 1)$ -ten Verlauf sind lediglich die Beobachtungen der ersten  $L$  Zeitpunkte ( $L < N$ )  $t_1, \dots, t_L$  bekannt. Diese werden in dem  $(L \times 1)$ -Beobachtungsvektor  $\mathbf{Y}_0 := (Y_{01}, \dots, Y_{0L})'$  zusammengefasst. Sei  $\mathbf{X}_0 \in \mathbb{R}^{L \times k}$  die Designmatrix,  $\boldsymbol{\beta}_0$  der  $(k \times 1)$ -Koeffizientenvektor und  $\boldsymbol{\epsilon}_0$  der  $(L \times 1)$ -Vektor der Störgrößen des neuen Verlaufs, sodass der Zusammenhang  $\mathbf{Y}_0 = \mathbf{X}_0 \boldsymbol{\beta}_0 + \boldsymbol{\epsilon}_0$  erfüllt ist. Dieser neue Verlauf folge ebenfalls den für das RCR-Modell gemachten Annahmen.

Das Ziel ist, eine Beobachtung  $Y_{0T} = \mathbf{x}'_T \boldsymbol{\beta}_0 + \epsilon_{0T}$  dieses neuen Verlaufs zu einem zukünftigen Zeitpunkt  $T$ ,  $T > t_L$ , zu prognostizieren. Rao empfiehlt in diesem Fall, die Beobachtung  $Y_{0T}$  mithilfe des Prädiktors aus Situation B vorherzusagen (Rao, 1975):

$$\hat{Y}_{0T} = \mathbf{x}'_T \left( \hat{\mathbf{C}} \hat{\boldsymbol{\beta}}_{0, \text{OLS}} + (\mathbf{I}_k - \hat{\mathbf{C}}) \hat{\mathbf{b}}_{\text{GLS}} \right),$$

wobei  $\hat{\mathbf{C}} := \hat{\boldsymbol{\Sigma}} \left( \hat{\boldsymbol{\Sigma}} + \hat{\sigma}_{\text{RCR}}^2 (\mathbf{X}'\mathbf{X})^{-1} \right)^{-1}$  ist. Die Schätzer  $\hat{\boldsymbol{\Sigma}}$ ,  $\hat{\sigma}_{\text{RCR}}^2$  und  $\hat{\mathbf{b}}_{\text{GLS}}$  werden hier allein auf Basis der Information der ersten  $I$  Verläufe  $\mathbf{Y}_1, \dots, \mathbf{Y}_I$  – d.h. ohne die Information des des neuen Verlaufs  $\mathbf{Y}_0$  – berechnet. Die Information des neuen Verlaufs fließt allein über den OLS-Schätzer  $\hat{\boldsymbol{\beta}}_{0, \text{OLS}} := (\mathbf{X}_0' \mathbf{X}_0)^{-1} \mathbf{X}_0' \mathbf{Y}_0$  in den Prädiktor  $\hat{Y}_{0T}$  ein. In Simulationsstudien konnte gezeigt werden, dass der Prädiktor  $\hat{Y}_{0T}$  anderen Prädiktoren für diese Situation überlegen ist (Bondeson, 1990).

Hinweis:

In diesem Auszug der Bachelorarbeit sind die Seiten 44 – 93 aus Datenschutzgründen nicht  
enthalten.

## Literaturverzeichnis

- Bondeson, J. (1990). *Prediction in random coefficient regression models*. Biometrical Journal, Vol. 32, No. 4, S. 387-405.
- Carter, R.L. und Yang, M.C.K. (1986). *Large sample inference in random coefficient regression models*. Communications in Statistics – Theory and Methods, Vol. 15, No. 8, S. 2507-2525.
- Casparry, W. und Wichmann, K. (1994). *Lineare Modelle*. Oldenbourg-Verlag, München.
- Frees, E.W. (2006). *Longitudinal and panel data – Analysis and applications in the social sciences*. Cambridge University Press, New York.
- Geberth, S. und Nowack, R. (2011). *Praxis der Dialyse*. Springer-Verlag, Berlin.
- Greene, W.H. (2008). *Econometric analysis*. 6. Auflage, Pearson-Verlag, New Jersey.
- Gumpertz, M. und Pantula, S.G. (1989). *A simple approach to inference in random coefficient models*. The American Statistician, Vol. 43, No. 4, S. 203-210.
- Harville, D. (1976). *Extension of the Gauss-Markov Theorem to include the estimation of random effects*. The Annals of Statistics, Vol. 4, No. 2, S. 384-395.
- Johansen, S. (1982). *Asymptotic inference in random coefficient regression models*. Scandinavian Journal of Statistics, Vol. 9, S. 201-207.
- Kadiyala, K.R. und Oberhelman, D. (1982). *Response predictions in regressions on panel data*. Communications in Statistics – Theory and Methods, Vol. 11, No. 23, S. 2699-2714.
- Laird, N.M. und Ware, J.H. (1982). *Random-effects models for longitudinal data*. Biometrics, Vol. 38, No. 4, S. 963-974.
- Patel, J.K. (1989). *Prediction intervals – A review*. Communications in Statistics – Theory and Methods, Vol. 18, No. 7, S. 2393-2465.
- Poi, B.P. (2003). *From the help desk: Swamy's random coefficients model*. The Stata Journal, Vol. 3, No. 3, S. 302-308.



- R Development Core Team (2010). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Wien.
- Rao, C.R. (1965). *The theory of least squares when the parameters are stochastic and its application to the analysis of growth curves*. *Biometrika*, Vol. 52, No. 3/4, S. 447-458.
- Rao, C.R. (1973). *Linear statistical inference and its applications*. 2. Auflage, Verlag John Wiley & Sons, New York.
- Rao, C.R. (1975). *Simultaneous estimation of parameters in different linear models and applications to biometric problems*. *Biometrics*, Vol. 31, No. 2, S. 545-554.
- Robinson, G.K. (1991). *That BLUP is a good thing: The estimation of random effects*. *Statistical Science*, Vol. 6, No. 1, S. 15-51.
- Schreiber, C., Al-Chalabi, A., Tanase, O. und Kreymann, B. (2009). *Grundlagen der Nieren- und Leberdialyse*. In: *Medizintechnik – Life science engineering*, S. 1519-1584, Herausgeber: Wintermantel, E. und Ha, S.W., Springer-Verlag, Berlin.
- Seber, G.A.F. und Lee, A.J. (2003). *Linear regression analysis*. 2. Auflage, Verlag John Wiley & Sons, New York.
- Seber, G.A.F. (2007). *A matrix handbook for statisticians*. Verlag John Wiley & Sons, New York.
- Swamy, P.A.V.B. (1971). *Statistical inference in random coefficient regression models*. Springer-Verlag, New York.
- Toutenburg, H. (1992). *Lineare Modelle*. Physica-Verlag, Heidelberg.
- Uhlin, F., Fridolin, I., Lindberg, L.G. und Magnusson, M. (2005). *Estimating total urea removal and protein catabolic rate by monitoring UV absorbance in spent dialysate*. *Nephrology Dialysis Transplantation*, Vol. 20, No. 7, S. 2458-2464.
- Uhlin, F., Fridolin, I., Lindberg, L.G. und Magnusson, M. (2006). *Dialysis dose (Kt/V) and clearance variation sensitivity using measurement of ultraviolet-absorbance (on-line), blood urea, dialysate urea and ionic dialysance*. *Nephrology Dialysis Transplantation*, Vol. 21, No. 8, S. 2225-2231.
- Urban, D. und Mayerl, J. (2006). *Regressionsanalyse: Theorie, Technik und Anwendung*. 2. und überarbeitete Auflage, VS Verlag für Sozialwissenschaften, Wiesbaden.

Verbeke, G. und Molenberghs, G. (2000). *Linear mixed models for longitudinal data*. Springer-Verlag, New York.

Weiss, R.E. (2005). *Modeling longitudinal data*. Springer-Verlag, New York.

Werner, J. (1992). *Numerische Mathematik 1 – Lineare und nichtlineare Gleichungssysteme, Interpolation, numerische Integration*. Vieweg-Verlag, Braunschweig.

Hinweis:

In diesem Auszug der Bachelorarbeit ist die Seite 97 aus Datenschutzgründen nicht enthalten.