

**Iterative Identifikation von Lokations- und Dispersionsfaktoren
bei nichtwiederholten Plänen**

Dissertation

zur Erlangung des Grades
einer Doktorin der Naturwissenschaften
der Technischen Universität Dortmund

Der Fakultät Statistik
der Technischen Universität Dortmund

vorgelegt von

Corinna Auer

Dortmund, August 2010

Inhaltsverzeichnis

1	Einleitung	1
2	Transformationsmodelle und GLM	5
2.1	Transformationsmodelle	6
2.1.1	Box-Cox Verfahren	8
2.1.2	Transformationsmodelle mit Dispersionsparameter	9
2.2	GLMs für Modelle mit Dispersionsfaktoren	12
2.2.1	Klassische Generalisierte Lineare Modelle	12
2.2.2	Gütemaße und Residuen	20
2.2.3	Quasi-Likelihood	24
2.2.4	Extended-Quasi-Likelihood und Pseudo-Likelihood	27

2.2.5	Restringierte EQL- und PL-Versionen	34
2.2.6	GLM basierte Methode zur Schätzung von Dispersionseffekten und Transformationsparametern	39
2.3	Erweiterung der Box-Cox-Methode	42
3	Asymptotische Eigenschaften	47
3.1	EQL und restringierte Versionen	49
3.1.1	Extended-Quasi-Likelihood	49
3.1.2	Restringierte EQL-Versionen	50
3.2	Restringierte PL-Version nach McCullagh und Nelder	51
3.3	Restringierte PL-Version nach Lee und Nelder	61
4	Effektidentifikation (Lokation und Dispersion)	69
4.1	Identifikation nach Pan und Taam	73
4.2	Überlegungen zur Devianzdifferenz	76
4.3	Identifikation mit Devianzdifferenzen	86
4.4	Vergleich der iterativen Verfahren	88
4.5	Beispiel	98

5 Zusammenfassung und Ausblick	101
A Regularitätsbedingungen	105
A.1 Alternative Regularitätsbedingung	105
A.2 Nachweis der Regularitätsbedingungen für PL^L	107
B Grenzwertbestimmungen für $\lambda \rightarrow 0$	117
C Tabellen und Grafiken	123
C.1 Tabellen	123
C.2 Grafiken	125
Literaturverzeichnis	133

Abkürzungen

$D = \mathcal{D} $	Anzahl der Dispersionseffekte
\mathcal{D}	Menge der Dispersionseffekte
Dev	Devianz
$\mathbf{e} = (e_1, \dots, e_n)^T$	Fehlervektor
EQL	Extended-Quasi-Likelihood
EQL^L	restringierte Extended-Quasi-Likelihood nach Lee und Nelder
EQL^M	restringierte Extended-Quasi-Likelihood nach McCullagh und Nelder
f_y	Dichte von y
$g(\cdot)$	Linkfunktion des Lokationsmodells
$h(\cdot)$	Linkfunktion des Dispersionsmodells
ℓ	Log-Likelihood
$L = \mathcal{L} $	Anzahl der Lokationseffekte
\mathcal{L}	Menge der Lokationseffekte
PL	Pseudo-Likelihood
PL^L	restringierte Pseudo-Likelihood nach Lee und Nelder
PL^M	restringierte Pseudo-Likelihood nach McCullagh und Nelder
QL	Quasi-Likelihood
QPL^L	restringierte Quasi-Pseudo-Likelihood nach Lee und Nelder
QPL^M	restringierte Quasi-Pseudo-Likelihood nach McCullagh und Nelder
r_P	Pearson-Residuen

r_D	Devianz Residuen
$s = L + D + 4$	Länge des Parametervektors $\theta =$ Anzahl zu schätzender Parameter
$V(\cdot)$	Varianzfunktion
$\mathbf{y} = (y_1, \dots, y_n)^T$	Vektor der Beobachtungen
$\mathbf{X}^{n \times L+1}$	Designmatrix des Lokationsmodell
$\mathbf{x}_i = (x_{i1}, \dots, x_{iL})^T$	Einstellungen der L Lokationseffekte bei Lauf $i, i = 1, \dots, n$
$\mathbf{x}_p = (x_{1p}, \dots, x_{np})^T$	Einstellungen von Lokationseffekt p in den n Läufen
$\mathbf{Z}^{n \times D+1}$	Designmatrix des Dispersionsmodells
$\mathbf{z}_i = (z_{i1}, \dots, z_{iD})^T$	Einstellung der D Dispersionseffekte bei Lauf $i, i = 1, \dots, n$
$\mathbf{z}_q = (z_{1q}, \dots, z_{nq})^T$	Einstellungen von Dispersionseffekt q in den n Läufen
$\alpha = (\alpha_0, \alpha_1, \dots, \alpha_L)^T$	Parametervektor der L Lokationseffekte
β	Parameter der Varianzfunktion
$\gamma = (\gamma_0, \gamma_1, \dots, \gamma_D)^T$	Parametervektor der D Dispersionseffekte
$\eta = \mathbf{X}\alpha$	linearer Prädiktor des Lokationsmodells
$\eta_i = \mathbf{x}_i^T \alpha$	linearer Prädiktor des Lokationsmodells in Lauf i
ϑ	natürlicher Parameter der Exponentialfamilie
$\theta = (\theta_1, \dots, \theta_s)^T$	Parametervektor
λ	Parameter der Linkfunktion
$\mu = (\mu_1, \dots, \mu_n)$	Erwartungswertvektor
$\xi = \mathbf{Z}\gamma$	linearer Prädiktor des Dispersionsmodells
$\xi_i = \mathbf{z}_i^T \gamma$	linearer Prädiktor des Dispersionsmodells in Lauf i
φ	Dispersionsparameter

Kapitel 1

Einleitung

In allen Anwendungsgebieten der Statistik gibt es nicht nur Größen, die den Erwartungswert einer Zielvariablen beeinflussen, so genannte Lokationseffekte, sondern auch Größen, die die Varianz beeinflussen, so genannte Dispersionseffekte. In einigen Gebieten reicht es aus, das Versuchsdesign so aufzustellen, dass es trotz der Varianz beeinflussenden Größen, eine große externe Validität gewährleistet, siehe Richter et al. (2010).

In technischen Anwendungen hat es sich jedoch seit der Einführung von Taguchi in den 50er Jahren des letzten Jahrhunderts etabliert, die Prozesse nicht nur bezüglich ihres Erwartungswertes sondern auch bezüglich ihrer Varianz zu optimieren. Die ursprünglichen Taguchi-Methoden beruhen dabei auf den so genannten Produktplänen. Sie basieren auf Wiederholungen der Designmatrix für das Lokationsmodell und erlauben neben der Schätzung der Lokationseffekte auch eine der Dispersionseffekte. Zusätzlich wird in technischen Prozessen häufig ein funktionaler Zusammenhang zwischen dem Erwartungswert

und der Varianz beobachtet. Durch geeignete Transformationen wird versucht, diesen Zusammenhang aufzuheben.

Alternativ zu den ursprünglichen Produktplänen, die auf Wiederholungen beruhen und somit eine Schätzung der Varianz erlauben, werden häufig so genannte Combined Arrays verwendet, vergleiche Steinberg und Bursztyn (1994). Sie werden häufig bevorzugt, da sie bei gleicher Designgröße eine bessere Vermischungsstruktur bezüglich der Lokationsfaktoren bieten und das Modell ein größeres Verständnis über die Quelle der Varianz liefert. Ein Nachteil der Combined Arrays ist jedoch, dass mit den ursprünglichen Auswertungsmethoden die Heteroskedastizität nur durch Wechselwirkungen zwischen Design- und Störgrößen erklärt werden kann. Effekte, die einen direkten Einfluss auf die Varianz ausüben, können nicht erkannt werden. Einen ausführlichen Vergleich der beiden Designs am Beispiel einer konkreten Anwendung wird in dem Paper von Kunert et al. (2007) erbracht. Außerdem können die für Produktpläne bestehenden Standardmethoden zur Bestimmung einer geeigneten Transformation auf Grund der fehlenden Wiederholungen nicht auf Combined Arrays übertragen werden.

Erdbrügge und Auer (2010) schlagen zwei Verfahren vor, die auch für nichtwiederholte Pläne sowohl Lokations- als auch Dispersionseffekte schätzen und zusätzlich einen geeigneten Transformationsparameter identifizieren können. Das eine Verfahren ist eine Verallgemeinerung der Box-Cox-Methode, siehe Box und Cox (1964), das andere beruht auf generalisierten linearen Modellen, wobei für dieses Verfahren Konsistenz, asymptotische Effizienz und Normalverteilung nachgewiesen werden konnte, wenn dabei die Parameter mit Hilfe der Pseudo-Likelihood geschätzt werden.

Ein weiterer Aspekt in der technischen Versuchsplanung ist die Optimierung der Prozesse. Auer et al. (2004) vergleichen verschiedene Möglichkeiten der multivariaten Optimierung bei Vorliegen von Dispersionseffekten anhand des Drückprozesses.

Der Augenmerk dieser Arbeit richtet sich auf die Identifikation aktiver Lokations- und Dispersionseffekte. Die Identifikation soll auf Basis von Devianzdifferenzen durchgeführt werden. Deshalb werden in dieser Arbeit auch die theoretischen Eigenschaften der durch die Minimierung der Devianz gewonnenen Schätzer untersucht werden. Die Arbeit gliedert sich wie folgt:

In Kapitel 2 werden beide von Erdbrügge und Auer (2010) vorgeschlagenen Verfahren noch einmal ausführlich beschrieben. Dazu wird vorab ein Überblick über bekannte Methoden zur Bestimmung eines Transformationsparameters gegeben und die Theorie der Generalisierten Linearen Modelle mit ihren Erweiterungen zur Berücksichtigung von Dispersionseffekten ausführlich erläutert. Im Zuge dessen werden auch restringierte Varianten der Pseudo-Likelihood vorgeschlagen.

Im darauf folgenden Kapitel 3 werden dann die theoretischen Eigenschaften des auf generalisierten linearen Modellen basierenden Verfahrens untersucht, wenn bei diesem statt der Pseudo-Likelihood die Extend-Quasi-Likelihood oder restringierte erweiterte Likelihood Varianten verwendet werden. Hierbei wird sich an dem von Erdbrügge und Auer (2010) geführten Beweis orientiert. Da den betrachteten Erweiterungen der Log-Likelihood keine echte Dichte zu Grunde liegt, wird dabei auf so genannte Quasi-Verteilungen zurückgegriffen, die zumindest für die restringierten Varianten der Pseudo-Likelihood explizit darstellbar sind. Basierend auf diesen Quasi-Verteilungen werden dann die theoretischen

Eigenschaften der Schätzer untersucht und, wenn möglich, auf die ursprünglichen Schätzer übertragen.

In Kapitel 4 wird eine auf den Devianzdifferenzen basierende Methode zur Identifizierung von aktiven Lokations- und Dispersionsfaktoren eingeführt. Diese wird mit der Methode von Pan und Taam (2002) verglichen. Simulationen zeigen, dass die vorgeschlagene Methode für die Identifikation aktiver Dispersionseffekte ähnlich gute Ergebnisse liefert, wie die von Pan und Taam. Das auf den Devianzdifferenzen basierende Verfahren ist dabei klarer beschrieben, benötigt keine normalverteilten Daten und besitzt nicht die Gefahr zu großer Modelle, da es auf einer Vorwärts- statt einer Rückwärtsselektion beruht.

Zum Schluss wird in Kapitel 5 eine Zusammenfassung der Ergebnisse und ein Ausblick auf mögliche weitere Untersuchungen gegeben.

Kapitel 2

Transformationsmodelle und generalisierte lineare Modelle

Hängt der Erwartungswert einer Zielgröße y von L unabhängigen Einflussgrößen ab, ist es wünschenswert, diese Abhängigkeit durch einen funktionalen Zusammenhang zu beschreiben. Häufig wird ein linearer Zusammenhang unterstellt, der mit Hilfe von unbekanntem, zu schätzenden Parametern beschrieben wird. In Matrixschreibweise lässt sich so ein linearer Zusammenhang in der Form $\mathbf{y} = \mathbf{X}\alpha + \mathbf{e}$ darstellen. Dabei bezeichnet $\mathbf{y} = (y_1, \dots, y_n)^\top$ den Vektor der n Realisationen der Zielvariablen. Die Designmatrix $\mathbf{X} \in \mathbb{R}^{n \times L+1}$ enthält in den $L + 1$ Spalten die zugehörigen Einstellungen $\mathbf{x}_1, \dots, \mathbf{x}_L$ der L Einflussgrößen in den n Läufen und eine Einserspalte $\mathbf{1}_n$ für das allgemeine Mittel. Die Menge der L Einflussgrößen wird im Folgenden mit \mathcal{L} bezeichnet und Menge der Lokationseffekte genannt. $\alpha = (\alpha_0, \alpha_1, \dots, \alpha_L)^\top$ gibt den zugehörigen Parametervektor einschließlich des allgemeinen Mittels wieder. Für den Fehlervektor $\mathbf{e} = (e_1, \dots, e_n)^\top$ wird im Allgemeinen die Annahme

getroffen, dass er Realisationen von stochastisch unabhängigen, normalverteilten Zufallsvariablen mit Erwartungswert Null und konstanter, erwartungswertunabhängiger Varianz σ^2 darstellt. Zusätzlich spiegelt das Modell die Annahme wieder, dass Additivität der Effekte vorliegt. Im Folgenden wird immer davon ausgegangen, dass die Matrix X vollen Rang besitzt und somit $(X^T X)^{-1}$ existiert.

2.1 Transformationsmodelle

Die oben beschriebenen einfachen linearen Modelle setzen voraus, dass die Fehler eine konstante Varianz besitzen. In technischen Anwendungen ist jedoch häufig zu beobachten, dass die Varianz mit steigendem Erwartungswert wächst, dass also ein funktionaler Zusammenhang zwischen diesen zwei Größen besteht. Um trotzdem mit obigem Modell arbeiten zu können, wird versucht, eine Transformation zu finden, die zu konstanter Varianz der transformierten Daten führt. Allgemein wird angenommen, dass ein Zusammenhang folgender Art besteht

$$\text{Var}(y_i) = \sigma_i^2 = v(\mu_i) \text{ und } E(y_i) = \mu_i > 0, i = 1, \dots, n.$$

Nun wird mit Hilfe der Delta-Methode, siehe Schervish (1995), S. 401f., eine Transformation $T(\cdot)$ bestimmt, so dass die transformierten Daten ungefähr eine konstante Varianz besitzen. Dabei wird die Tatsache ausgenutzt, dass sich die transformierten Daten mit einer Taylorentwicklung um μ_i approximativ wie folgt darstellen lassen:

$$T(y_i) \approx T(\mu_i) + (y_i - \mu_i) \frac{\partial}{\partial y_i} T(y_i) \Big|_{y_i = \mu_i}.$$

Da die transformierten Daten eine konstante Varianz besitzen sollen, muss gelten:

$$\begin{aligned}
 c &= \text{Var}(T(y_i)) \\
 &\approx \text{Var}\left(T(\mu_i) + (y_i - \mu_i) \frac{\partial}{\partial y_i} T(y_i) \Big|_{y_i=\mu_i}\right) \\
 &= \text{Var}(y_i) \left(\frac{\partial}{\partial y_i} T(y_i) \Big|_{y_i=\mu_i}\right)^2 \\
 &= v(\mu_i) \left(\frac{\partial}{\partial y_i} T(y_i) \Big|_{y_i=\mu_i}\right)^2.
 \end{aligned}$$

Somit muss für die Transformationsvorschrift $T(\cdot)$ gelten: $T(\mu_i) = \int \frac{c}{\sqrt{v(\mu_i)}} d\mu_i$. In dieser Arbeit wird immer von dem häufig unterstellten exponentiellen Zusammenhang folgender Art ausgegangen:

$$\sigma_i^2 = \varphi \mu_i^{2\beta}, \quad i = 1, \dots, n, \quad \text{und } \varphi > 0. \quad (2.1)$$

Hierbei resultiert die Delta-Methode durch Bildung der zugehörigen Stammfunktion in einer so genannten Power-Transformation:

$$T(y_i) \propto \begin{cases} \frac{y_i^{1-\beta} - 1}{1-\beta} & : \beta \neq 1 \\ \log y_i & : \beta = 1 \end{cases}. \quad (2.2)$$

Hierbei wird angenommen, dass $y_i > 0$ für alle $i, i = 1, \dots, n$. Ist diese Annahme nicht erfüllt, wird allen y_i eine Konstante hinzu addiert.

Im Allgemeinen wird zwar von einem Zusammenhang der Art (2.1) ausgegangen, der genaue Wert für den Parameter β ist jedoch meist unbekannt und muss geschätzt werden. Ein bekanntes Verfahren zur Schätzung des Transformationsparameters β ist die β -Methode von Logothetis (1988), die für wiederholte Pläne geeignet ist. Hierbei werden

für jede Einstellung der Einflussgrößen sowohl der Mittelwert als auch die Standardabweichung berechnet. Trägt man die logarithmierten Standardabweichungen gegen die logarithmierten Mittelwerte ab, so entspricht die Steigung dem Transformationsparameter β und kann mit einem einfachen linearen Modell geschätzt werden.

2.1.1 Box-Cox Verfahren

Box und Cox (1964) benutzen zur Bestimmung des Transformationsparameters $\lambda = 1 - \beta$ die Maximum-Likelihood Methode. Dazu transformieren sie die Beobachtungen gemäß einer zu (2.2) proportionalen Transformationsvorschrift $T_\lambda(\mathbf{y})$ und unterstellen für die transformierten Daten Normalverteilung und konstante Varianz, das heißt sie gehen von folgendem Modell aus:

$$T_\lambda(\mathbf{y}) = \mathbf{X}\alpha + \mathbf{e} \quad \text{mit} \quad \mathbf{e} \sim N(\mathbf{0}, \sigma^2 I_n) \quad \text{und} \quad T_\lambda(y_i) = \begin{cases} \frac{y_i^\lambda - 1}{\lambda} & : \lambda \neq 0 \\ \log y_i & : \lambda = 0 \end{cases}. \quad (2.3)$$

Die Dichte der untransformierten Zielvariable \mathbf{y} kann mit Hilfe des Transformationssatzes für stetige Dichten bestimmt werden zu

$$f(\mathbf{y} | \alpha, \sigma^2, \lambda) = f(T_\lambda(\mathbf{y} | \alpha, \sigma^2, \lambda)) \det J,$$

wobei $\det J = (\prod_{i=1}^n y_i)^{\lambda-1}$ die zugehörige Jacobideterminante ist. Die Maximierung der resultierenden Log-Likelihood bezüglich der Parameter α, σ^2 und λ führt zu folgenden expliziten Lösungen für $\hat{\alpha}$ bzw. $\hat{\sigma}^2$:

$$\hat{\alpha} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top T_\lambda(\mathbf{y}) \quad \text{und} \quad \hat{\sigma}^2 = \frac{1}{n} T_\lambda(\mathbf{y})^\top \omega^{-1}(X) T_\lambda(\mathbf{y}),$$

wobei $\omega^\perp(X) = I_n - X(X^\top X)^{-1}X^\top$ die Orthogonalprojektion auf X ist. Der Schätzer für λ kann nur numerisch durch Maximierung der Log-Likelihood

$$\ell\left(\lambda \mid \hat{\alpha}, \hat{\sigma}^2, \mathbf{y}\right) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log\left(\frac{1}{n} T_\lambda(\mathbf{y})^\top \omega^\perp(X) T_\lambda(\mathbf{y})\right) - \frac{n}{2} + (\lambda - 1) \mathbf{1}_n^\top \log \mathbf{y}$$

gelöst werden. Die von Box und Cox (1964) vorgeschlagene standardisierte Transformationsvorschrift

$$\tilde{T}_\lambda(\mathbf{y}) = \frac{T_\lambda(\mathbf{y})}{\dot{\mathbf{y}}^{\lambda-1}},$$

wobei $\dot{\mathbf{y}} = \left(\prod_{i=1}^n y_i\right)^{\frac{1}{n}}$ das geometrische Mittel der Beobachtungen ist, führt zu folgender vereinfachter Log-Likelihood

$$\ell\left(\lambda \mid \hat{\alpha}, \hat{\sigma}^2, \mathbf{y}\right) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log\left(\frac{1}{n} \underbrace{\tilde{T}_\lambda(\mathbf{y})^\top \omega^\perp(X) \tilde{T}_\lambda(\mathbf{y})}_{SSE(\lambda)}\right) - \frac{n}{2}.$$

Äquivalent zur Maximierung dieser Log-Likelihood kann nun die Fehlerquadratsumme $SSE(\lambda)$ bezüglich λ minimiert werden. In praktischen Anwendungen reicht es oft aus, $\hat{\lambda}$ grafisch zu bestimmen. Dazu wird die Fehlerquadratsumme $SSE(\lambda)$ gegen eine Sequenz $\Lambda = \{\lambda_1, \dots, \lambda_r\}$ abgetragen, und der Wert als Schätzer für λ genommen, der die Fehlerquadratsumme ungefähr minimiert. Häufig wird $-1, 0, \frac{1}{2}, 1$ oder 2 als Transformationsparameter gewählt, so dass der Kehrwert, der Logarithmus, die Wurzel, die Originalwerte oder die Quadrate der Beobachtungen analysiert werden.

2.1.2 Transformationsmodelle mit Dispersionsparameter

Neben der Möglichkeit, dass Faktoren einen indirekten Einfluss auf die Varianz ausüben, nämlich über einen funktionellen Zusammenhang zwischen Erwartungswert und Varianz,

ist es auch möglich, dass es Faktoren gibt, die einen direkten Einfluss auf die Varianz ausüben, so genannte Dispersionsfaktoren. Dem kann Rechnung getragen werden, indem die Konstante φ aus Formel (2.1) nun durch ein φ_i in Abhängigkeit von Dispersionsfaktoren modelliert wird. Hierbei wird angenommen, dass der Zusammenhang folgendermaßen aussieht:

$$\sigma_i^2 = \mu_i^{2\beta} \exp\{\mathbf{z}_i^\top \boldsymbol{\gamma}\}, \quad i = 1, \dots, n. \quad (2.4)$$

Dabei stellt $\mathbf{z}_i = (1, z_{i1}, \dots, z_{iD})^\top$ den Vektor der Einstellungen für das allgemeine Mittel und die D Dispersionseffekte in Lauf $i, i = 1, \dots, n$, dar. Die Menge der D Dispersionseffekte wird im Folgenden mit \mathcal{D} bezeichnet. Dabei muss die Menge der Dispersionseffekte \mathcal{D} nicht notwendigerweise disjunkt zur Menge der Lokationseffekte \mathcal{L} sein. $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \dots, \gamma_D)^\top$ gibt den zugehörigen Parametervektor wieder. Der Gebrauch der Exponentialfunktion ist üblich, da er positive Schätzer für die Varianz garantiert und eine einfache Interpretation ermöglicht. Existiert kein funktioneller Zusammenhang, so reduziert sich σ_i^2 zu $\exp\{\mathbf{z}_i^\top \boldsymbol{\gamma}\}$ für alle $i, i = 1, \dots, n$. Dieses Modell entspricht der von Aitken (1987) und Engel und Huele (1996) vorgeschlagenen multiplikativen Heteroskedastizität. Liegt Varianzmodell (2.4) vor, so sind die aus der β -Methode erhaltenen Schätzer inkonsistent. Existiert jedoch höchstens ein Dispersionseffekt, so kann die generalisierte β -Methode von Kunert und Lehmkuhl (1998) zur konsistenten Schätzung herbeigezogen werden. Hierbei werden für die verschiedenen Einstellungen des Dispersionsfaktors parallele Geraden angepasst.

Eine weitere Möglichkeit, den Transformationsparameter bei Vorliegen von Dispersionseffekten zu schätzen, ist der von Box (1988) vorgeschlagene λ -Plot. Er nimmt multiplikative

Heteroskedastizität für die transformierten Daten $T_\lambda(\mathbf{y})$ an, d.h. er unterstellt

$$T_\lambda(\mathbf{y}) = \mathbf{X}\alpha + \mathbf{e}, \text{ wobei } \mathbf{e} \sim N(\mathbf{0}, \Sigma) \text{ und } \Sigma = \text{diag}(\sigma_i^2) = \text{diag}(\exp\{\mathbf{z}_i^\top \boldsymbol{\gamma}\}) \quad (2.5)$$

und schätzt für eine Sequenz $\Lambda = \{\lambda_1, \dots, \lambda_r\}$ standardisierte Lokations- und Dispersionseffekte. Diese Schätzer werden in zwei separaten Grafiken abgetragen. Der Transformationsparameter wird nun so gewählt, dass möglichst wenige Dispersionseffekte identifiziert werden, gleichzeitig aber mehrere klare Lokationseffekte und möglichst wenige Wechselwirkungen sichtbar sind.

Weder die β -Methoden (die ursprüngliche sowie die generalisierte) noch der λ -Plot können jedoch bei Vorliegen von nichtwiederholten Plänen angewandt werden.

Erdbrügge und Auer (2010) schlagen zwei Verfahren vor, welche auch im Fall von nichtwiederholten Plänen den Transformationsparameter $\lambda = 1 - \beta$ sowie die Parametervektoren der Lokations- und Dispersionseffekte gleichzeitig schätzen können. Das eine Verfahren ist eine Erweiterung des oben erläuterten Box-Cox-Verfahrens, das andere basiert auf generalisierten linearen Modellen. Beide Erweiterungen werden in den folgenden Abschnitten ausführlich erläutert.

2.2 Generalisierte Lineare Modelle für Modelle mit Dispersionsfaktoren

2.2.1 Klassische Generalisierte Lineare Modelle

Mit Hilfe der obigen Transformationsmodelle sollen die Annahmen der linearen Modelle, d.h. Normalverteilung der Fehler, Additivität der Effekte und eine vom Erwartungswert unabhängige konstante Varianz erzielt werden. Es ist jedoch nicht immer möglich, eine einzige Transformation zu finden, die zur gleichzeitigen Erfüllung aller Annahmen führt. Nelder und Lee (1991) führen das Beispiel von Poisson-verteilten Zufallsvariablen y_1, \dots, y_n mit Erwartungswerten μ_1, \dots, μ_n auf. Um approximative Normalverteilung zu erreichen, wird die Transformation $z = y^{2/3}$ benötigt, konstante Varianzen hingegen werden mit $z = y^{1/2}$ erzielt und für die Additivität der Effekte wird meist eine Log-Transformation benötigt. Eine Möglichkeit, dieses Problem zu umgehen, bieten generalisierte lineare Modelle (GLM), welche die Daten in der Originalskala modellieren. Sie arbeiten mit dem Konzept der Exponentialfamilie, die wie folgt definiert ist (siehe Fahrmeir und Tutz (1997), Anhang A.1):

Definition 1 (*Einparametrische Exponentialfamilie mit Dispersionsparameter*)

Sei der statistische Raum $(\Omega, \mathcal{A}, \mathcal{P})$ mit Parametrisierung $\mathcal{P} = \{P_\vartheta : \vartheta \in \Theta\}$ und die Zufallsvariable y gegeben. Die Familie von Verteilungen $\mathcal{P}^y = \{P_\vartheta^y : \vartheta \in \Theta\}$ auf $(\mathbb{R}, \mathcal{B})$ heißt einparametrische Exponentialfamilie mit Dispersionsparameter φ , falls folgendes gilt:

1. $\vartheta \in \Theta \subset \mathbb{R}$
2. \exists ein σ -endliches Maß μ auf $(\mathbb{R}, \mathcal{B})$, so dass für alle $\vartheta \in \Theta$ eine μ -Dichte f_{ϑ} von P_{ϑ}^y existiert
3. f_{ϑ} ist darstellbar als $f(y|\vartheta, \varphi) = \exp \left\{ \frac{y\vartheta - b(\vartheta)}{\varphi} + c(y, \varphi) \right\}$,
wobei $b : \Theta \rightarrow \mathbb{R}$ beliebig oft differenzierbar, $\varphi \in \mathbb{R}^+$ und $c : \mathbb{R} \times \mathbb{R}^+ \rightarrow \mathbb{R}$ messbar auf \mathcal{B} ist.

Dabei heißt ϑ natürlicher Parameter und φ Dispersionsparameter.

Satz 1

Für den Erwartungswert und die Varianz einer Zufallsvariablen y , deren Verteilung zu einer einparametrischen Exponentialfamilie mit Dispersionsparameter φ gehört, gilt:

$$E\{y|\vartheta\} = \mu(\vartheta) = \frac{\partial}{\partial \vartheta} b(\vartheta)$$

$$Var\{y|\vartheta\} = \varphi \frac{\partial^2}{\partial \vartheta^2} b(\vartheta) = \varphi \frac{\partial}{\partial \vartheta} \mu(\vartheta) > 0.$$

Beweis: Laut Lehmann (1986), S. 29, ist in Exponentialfamilien die Vertauschung von Integration und Differentiation erlaubt. Somit gilt:

$$\begin{aligned}
\text{(i)} \quad E \left\{ \frac{\partial}{\partial \vartheta} \log f(y | \vartheta, \varphi) \right\} &= \int \left\{ \frac{\partial}{\partial \vartheta} \log f(y | \vartheta, \varphi) \right\} f(y | \vartheta, \varphi) \, dy \\
&= \int \frac{\partial}{\partial \vartheta} f(y | \vartheta, \varphi) \, dy = \frac{\partial}{\partial \vartheta} \int f(y | \vartheta, \varphi) \, dy = \frac{\partial}{\partial \vartheta} 1 = 0 \quad \text{und} \\
\text{(ii)} \quad E \left\{ \frac{\partial^2}{\partial \vartheta^2} \log f(y | \vartheta, \varphi) \right\} + E \left\{ \left(\frac{\partial}{\partial \vartheta} \log f(y | \vartheta, \varphi) \right)^2 \right\} \\
&= \int \left[-\frac{1}{f(y | \vartheta, \varphi)^2} \left(\frac{\partial}{\partial \vartheta} f(y | \vartheta, \varphi) \right)^2 + \frac{1}{f(y | \vartheta, \varphi)} \frac{\partial^2}{\partial \vartheta^2} f(y | \vartheta, \varphi) \right] f(y | \vartheta, \varphi) \, dy \\
&\quad + \int \left[\frac{1}{f(y | \vartheta, \varphi)} \frac{\partial}{\partial \vartheta} f(y | \vartheta, \varphi) \right]^2 f(y | \vartheta, \varphi) \, dy \\
&= \int \frac{\partial^2}{\partial \vartheta^2} f(y | \vartheta, \varphi) \, dy \\
&= \frac{\partial^2}{\partial \vartheta^2} \int f(y | \vartheta, \varphi) \, dy = 0.
\end{aligned}$$

Mit

$$\frac{\partial}{\partial \vartheta} \log f(y | \vartheta, \varphi) = \frac{y - \frac{\partial}{\partial \vartheta} b(\vartheta)}{\varphi} \quad \text{und} \quad \frac{\partial^2}{\partial \vartheta^2} \log f(y | \vartheta, \varphi) = -\frac{1}{\varphi} \frac{\partial^2}{\partial \vartheta^2} b(\vartheta)$$

folgt für den Erwartungswert von y :

$$\begin{aligned}
E \left\{ \frac{\partial}{\partial \vartheta} \log f(y | \vartheta, \varphi) \right\} &= E \left\{ \frac{y - \frac{\partial}{\partial \vartheta} b(\vartheta)}{\varphi} \right\} = \frac{\mu(\vartheta) - \frac{\partial}{\partial \vartheta} b(\vartheta)}{\varphi} \stackrel{\text{(i)}}{=} 0 \\
\Leftrightarrow \mu(\vartheta) &= \frac{\partial}{\partial \vartheta} b(\vartheta).
\end{aligned}$$

Und für die Varianz von y ergibt sich:

$$\begin{aligned}
\text{Var} \{y | \vartheta, \varphi\} &= E \left\{ \left(y - \frac{\partial}{\partial \vartheta} b(\vartheta) \right)^2 \right\} = E \left\{ \varphi^2 \left[\frac{\partial}{\partial \vartheta} \log f(y | \vartheta) \right]^2 \right\} \\
&\stackrel{\text{(ii)}}{=} -\varphi^2 E \left\{ \frac{\partial^2}{\partial \vartheta^2} \log f(y | \vartheta) \right\} = -\varphi^2 E \left\{ -\frac{1}{\varphi} \frac{\partial^2}{\partial \vartheta^2} b(\vartheta) \right\} = \varphi \frac{\partial^2}{\partial \vartheta^2} b(\vartheta).
\end{aligned}$$

q.e.d.

Die Funktion $V : \mathbb{R} \rightarrow \mathbb{R}$ mit $V(\mu(\vartheta)) := \frac{\partial^2}{\partial \vartheta^2} b(\vartheta) = \frac{\partial}{\partial \vartheta} \mu(\vartheta)$, welche den Zusammenhang zwischen Varianz und Erwartungswert beschreibt, wird Varianzfunktion genannt, siehe McCullagh und Nelder (1989), S.28 f.

Beobachtungen, die einer Exponentialfamilie entstammen, können mit generalisierten linearen Modellen, die wie folgt definiert sind, in der Originalskala modelliert werden, siehe Fahrmeir und Tutz (1997), Anhang A.1:

Definition 2 (*Generalisiertes Lineares Modell (GLM)*)

Es seien die Zufallsvariablen y_1, \dots, y_n , die zugehörigen Vektoren $\mathbf{x}_i, i = 1, \dots, n$ mit den jeweiligen Einstellungen der L Lokationseffekte und der Dispersionsparameter φ gegeben.

Weiter gelte, dass $y_1 | (\mathbf{x}_1, \varphi), \dots, y_n | (\mathbf{x}_n, \varphi)$ stochastisch unabhängig sind. Ein generalisiertes lineares Modell ist dann durch eine Verteilungs- und eine Strukturannahme gegeben:

Verteilungsannahme:

Die Verteilung P^{y_i} gehört für alle $i, i = 1, \dots, n$, zu einer einparametrischen Exponentialfamilie mit natürlichem Parameter ϑ_i und Dispersionsparameter φ .

Strukturannahme:

Der Erwartungswert $\mu_i = E(y_i | \vartheta_i, \varphi)$ hängt über eine eindeutige Linkfunktion g von dem linearen Prädiktor $\eta_i = \mathbf{x}_i^\top \alpha$ wie folgt ab: $g(\mu_i(\vartheta_i)) = g\left(\frac{\partial}{\partial \vartheta_i} b(\vartheta_i)\right) = \eta_i(\alpha) = \mathbf{x}_i^\top \alpha$. Dabei ist $g : \mathbb{R} \rightarrow \mathbb{R}$ eine bijektive beliebig oft differenzierbare Funktion und $\frac{\partial b}{\partial \vartheta_i}(\cdot)$ bijektiv.

Dabei wird angenommen, dass eine Linkfunktion $g(\cdot)$ existiert, so dass die geforderte Additivität der Effekte gegeben ist. Der Zusammenhang zwischen Varianz und Erwartungswert

wird durch die Varianzfunktion $V(\mu)$ beschrieben, die genau wie der Dispersionsparameter φ durch die Exponentialfamilie festgelegt ist. Die Linkfunktion $g(\cdot) = \left(\frac{\partial b}{\partial \vartheta}\right)^{-1}(\cdot)$, heißt natürliche oder kanonische Linkfunktion und resultiert in $\vartheta_i = \mathbf{x}_i^\top \alpha$.

Da dem klassischen GLM-Ansatz eine echte Dichte zu Grunde liegt, kann die Log-Likelihood $\ell(\vartheta | \mathbf{y}, \varphi)$ bestimmt werden zu

$$\ell(\vartheta | \mathbf{y}, \varphi) = \sum_{i=1}^n \left[\frac{y_i \vartheta_i - b(\vartheta_i)}{\varphi} + \tilde{c}(y_i, \varphi) \right]. \quad (2.6)$$

Die Parameter in einem GLM, der natürliche Parameter ϑ , der Erwartungswert μ und der Parameter α des linearen Prädiktors, hängen voneinander ab. Unter Ausnutzung dieser Beziehungen kann die Log-Likelihood auch in Abhängigkeit von α oder μ dargestellt werden: Aus der Tatsache, dass $\mu_i(\alpha) = g^{-1}(\mathbf{x}_i^\top \alpha)$ und $\mu_i(\vartheta_i) = \frac{\partial b}{\partial \vartheta_i}(\vartheta_i)$, folgt

$$\vartheta_i = \left(\frac{\partial b}{\partial \vartheta_i}\right)^{-1}(\mu_i) = \left(\frac{\partial b}{\partial \vartheta_i}\right)^{-1}(g^{-1}(\mathbf{x}_i^\top \alpha)). \quad (2.7)$$

Einsetzen von (2.7) in (2.6) führt zu folgenden alternativen Schreibweisen der Log-Likelihood:

$$\ell(\mu | \mathbf{y}, \varphi) = \sum_{i=1}^n \left[\frac{y_i \left(\frac{\partial b}{\partial \vartheta_i}\right)^{-1}(\mu_i) - b\left[\left(\frac{\partial b}{\partial \vartheta_i}\right)^{-1}(\mu_i)\right]}{\varphi} + \tilde{c}(y_i, \varphi) \right] \quad (2.8)$$

bzw.

$$\ell(\alpha | \mathbf{y}, \varphi) = \sum_{i=1}^n \left[\frac{y_i \left(\frac{\partial b}{\partial \vartheta_i}\right)^{-1}(g^{-1}(\mathbf{x}_i^\top \alpha)) - b\left[\left(\frac{\partial b}{\partial \vartheta_i}\right)^{-1}(g^{-1}(\mathbf{x}_i^\top \alpha))\right]}{\varphi} + \tilde{c}(y_i, \varphi) \right] \quad (2.9)$$

Durch Lösen der so genannten Scoregleichungen

$$\begin{aligned}
\mathcal{S}(\alpha_p) &= \sum_{i=1}^n \frac{\partial \ell(\alpha_p | y_i, \varphi)}{\partial \alpha_p} = \sum_{i=1}^n \frac{\partial \ell(\vartheta_i | y_i, \varphi)}{\partial \alpha_p} \\
&= \sum_{i=1}^n \frac{\partial \ell(\vartheta_i | y_i, \varphi)}{\partial \vartheta_i} \frac{\partial \vartheta_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \alpha_p} \\
&= \sum_{i=1}^n \frac{y_i - \mu_i}{\varphi} \frac{1}{V(\mu_i)} \frac{\partial g^{-1}(\eta_i)}{\partial \eta_i} x_{ip} = 0, \quad p = 0, \dots, L.
\end{aligned}$$

können die ML-Schätzer für den Parametervektor $\alpha = (\alpha_0, \dots, \alpha_L)^\top$ der Lokationseffekte gewonnen werden.

Lehmann (1983) und Wasan (1970) geben die im Folgenden aufgeführten elementaren Regularitätsbedingungen an, unter denen die Konsistenz und Effizienz der resultierenden ML-Schätzer gezeigt werden kann. Der darauf folgende Satz stellt Bedingungen auf, unter denen die obigen wünschenswerten Eigenschaften der ML-Schätzer gegeben sind.

Bedingung 1 (*Elementare Regularitätsbedingungen*)

Es sei $\theta = (\theta_1, \dots, \theta_s)^\top \in \Theta \subset \mathbb{R}^s$ ein Parametervektor und $\mathbf{x}_i \in \mathcal{X}$ ein Vektor von Ausprägungen möglicher Einflussvariablen auf die Momente von y_i . Die Verteilung von y_i mit zugehöriger Dichte $f(y_i | \theta, \mathbf{x}_i)$ gegeben θ und \mathbf{x}_i bezüglich eines σ -endlichen Maßes ν werde mit P_θ bezeichnet. Weiter gelte:

(E1) Die Beobachtungen y_1, \dots, y_n sind stochastisch unabhängig.

(E2) $P_\theta \neq P_{\theta'}$ für $\theta \neq \theta'$ und $\theta, \theta' \in \Theta$.

(E3) Die Menge $\{y_i : f(y_i | \theta, \mathbf{x}_i) > 0\}$ ist unabhängig von θ , d. h. die Verteilungen P_θ haben einen gemeinsamen Träger.

Satz 2

Der Beobachtungsvektor \mathbf{y} sei gegeben durch $\mathbf{y} = (y_1, \dots, y_n)^\top$ und zu schätzen sei der wahre Parametervektor $\theta_0 = (\theta_{01}, \dots, \theta_{0s})^\top \in \Theta$. Es gelte die Bedingung 1 und zusätzlich die folgenden Regularitätsbedingungen:

(R1) Es gibt eine offene Teilmenge $\tilde{\Theta}$ von Θ , die den wahren Parameter θ_0 enthält, so dass für jedes $\theta \in \tilde{\Theta}$ alle Ableitungen

$$\frac{\partial^3}{\partial \theta_j \partial \theta_k \partial \theta_l} f(y_i | \theta, \mathbf{x}_i), \quad j, k, l \in \{1, \dots, s\}$$

der Dichtefunktion existieren (für fast alle y_i).

(R2) Die ersten und zweiten Ableitungen der Log-Likelihood $\ell(\theta | y_i) := \ell(\theta | y_i, \mathbf{x}_i)$ erfüllen die Gleichungen

$$\begin{aligned} (a) \quad E_\theta \left\{ \frac{\partial}{\partial \theta_j} \ell(\theta | y_i) \right\} &= 0 \quad \text{für } j = 1, \dots, s, \\ (b) \quad \mathcal{I}_{jk}(\theta) &= E_\theta \left\{ \frac{\partial}{\partial \theta_j} \ell(\theta | y_i) \cdot \frac{\partial}{\partial \theta_k} \ell(\theta | y_i) \right\} = E_\theta \left\{ -\frac{\partial^2}{\partial \theta_j \partial \theta_k} \ell(\theta | y_i) \right\} \\ &\quad \text{für } j, k = 1, \dots, s. \end{aligned}$$

(R3) Die $s \times s$ Matrix $\mathcal{I}(\theta)$ ist positiv definit und alle $\mathcal{I}_{jk}(\theta)$ sind endlich, so dass $\mathcal{I}(\theta)^{-1}$ existiert.

(R4) Es gibt Funktionen $H_{jkl}(y_i)$ (evtl. von θ_0 abhängig) mit $E_{\theta_0} \{H_{jkl}(y_i)\} = \tilde{H}_{jkl} < \infty$, so dass gilt:

$$\left| \frac{\partial^3}{\partial \theta_j \partial \theta_k \partial \theta_l} \log f(y_i | \theta, \mathbf{x}_i) \right| \leq H_{jkl}(y_i) \quad \forall \theta \in \tilde{\Theta} \quad \forall j, k, l.$$

Dann strebt für $n \rightarrow \infty$ die Wahrscheinlichkeit gegen 1, dass Lösungen $\hat{\theta} := \hat{\theta}(\mathbf{y})$ der Score-Gleichungen existieren, für die folgendes gilt:

- (i) Der Ausdruck $\sqrt{n}(\hat{\theta} - \theta_0)$ ist asymptotisch normalverteilt mit Erwartungswert $\mathbf{0}_s$ und Varianz-Kovarianz Matrix $[\mathcal{I}(\theta_0)]^{-1}$, wobei $\mathcal{I}(\theta_0)$ die Informationsmatrix nach Fisher ist.

Somit gilt insbesondere:

- (ii) Die Schätzer $\hat{\theta}_j$ sind konsistent für θ_{0j} .
- (iii) Die Schätzer $\hat{\theta}_j$ sind asymptotisch effizient, d. h.

$$\sqrt{n}(\hat{\theta}_j - \theta_{0j}) \xrightarrow{d} N(0, [\mathcal{I}(\theta_0)]_{jj}^{-1}) .$$

Beweis: siehe Lehmann (1983) und Wasan (1970).

Erdbrügge (2003) zeigt, dass für den Parametervektor α die Voraussetzungen für Satz 2 erfüllt sind, wenn die Inverse der Linkfunktion $g(\cdot)^{-1}$ dreimal stetig differenzierbar ist, $(\mathbf{X}^T \mathbf{X})^{-1}$ existiert und für $n \rightarrow \infty$ der Raum \mathcal{X} der Einstellungen der Einflussfaktoren beschränkt bleibt. Somit gibt es mit einer Wahrscheinlichkeit, die für $n \rightarrow \infty$ gegen 1 strebt, eine eindeutige Lösung der Scoregleichungen. Die daraus resultierenden Schätzer für $\theta = \alpha$ sind konsistent und asymptotisch effizient und normalverteilt.

Im Allgemeinen können die Scoregleichungen $\mathcal{S}(\theta) := \frac{\partial}{\partial \theta} \ell(\theta | \mathbf{y}) = \mathbf{0}_s$ jedoch nicht analytisch gelöst werden. Stattdessen wird meist das Newton-Verfahren mit Fisher-Scoring verwendet. Hierbei wird ein vorläufiger Schätzer θ_0 bestimmt. Dieser wird dann iterativ bis zum Erreichen von Konvergenz verbessert: In Schritt $k + 1$ wird der Schätzer $\theta^{(k+1)}$

gewonnen durch: $\theta^{(k+1)} = \theta^{(k)} - \left(\frac{\partial \mathcal{S}(\theta)}{\partial \theta} \Big|_{\theta=\theta^{(k)}} \right)^{-1} \mathcal{S}(\theta^{(k)})$. Beim Fisher-Scoring wird die Matrix der Ableitungen der Scorefunktion $\frac{\partial \mathcal{S}(\theta)}{\partial \theta} \Big|_{\theta=\theta^{(k)}}$ durch ihren Erwartungswert, die negative Fisher-Information $\mathcal{I}(\theta^{(k)})$ ersetzt, da diese leichter zu berechnen ist. Somit folgt $\theta^{(k+1)} = \theta^{(k)} + \mathcal{I}(\theta^{(k)})^{-1} \mathcal{S}(\theta^{(k)})$, siehe McCullagh und Nelder (1989), S.40.

In praktischen Fällen wird jedoch häufig beobachtet, dass die empirische Varianz größer oder auch kleiner ist als durch die Exponentialfamilie impliziert wird, es liegt Über- bzw. Unterdispersion vor. In diesen Fällen kann der Dispersionsparameter φ auch aus den Daten geschätzt werden. Ist α bekannt, so kann der erwartungstreue Momentenschätzer

$$\tilde{\varphi} = \frac{1}{n} \sum_{i=1}^n \frac{(y_i - g^{-1}(\mathbf{x}_i^T \alpha))^2}{V(g^{-1}(\mathbf{x}_i^T \alpha))}$$

verwendet werden. Wird α durch die Maximum-Likelihood Methode geschätzt, ist

$$\hat{\varphi} = \frac{1}{n - \text{rg}(X)} \sum_{i=1}^n \frac{(y_i - g^{-1}(\mathbf{x}_i^T \hat{\alpha}))^2}{V(g^{-1}(\mathbf{x}_i^T \hat{\alpha}))} \quad (2.10)$$

ein erwartungstreuer Schätzer für den Dispersionsparameter φ , siehe Firth (1991). Der Dispersionsparameter φ ist dabei unabhängig vom Laufindex i , kann also nicht wie der Erwartungswert μ_i über einen linearen Prädiktor modelliert werden, d.h. im klassischen generalisierten linearen Modell können keine Dispersionsfaktoren berücksichtigt werden.

2.2.2 Gütemaße und Residuen

Nach der Anpassung eines Modells mit p Parametern ist es sinnvoll, die Güte des Modells zu bestimmen. Im linearen Modell geschieht dies oft durch das Bestimmtheitsmaß. Dieses

ist hier jedoch nicht sinnvoll, weil es auf den Annahmen des linearen Modells aufbaut. Firth (1991) beschreibt zwei für GLMs häufig benutzte Gütemaße. Die Devianz vergleicht die Log-Likelihood $\ell(\hat{\mu} | \mathbf{y}, \varphi)$ des angepassten Modells mit der Log-Likelihood $\ell(\mathbf{y} | \mathbf{y}, \varphi)$ eines saturierten Modells, d.h. eines Modells, welches die Beobachtungen exakt wiedergeben würde. Sie ist wie folgt definiert:

Definition 3 ((skalierte) Devianz)

Gegeben seien die unabhängigen Beobachtungen $\mathbf{y} = (y_1, \dots, y_n)$ mit zugehöriger Log-Likelihood $\ell(\mu | \mathbf{y}, \varphi)$ und die durch ein GLM bestimmten Schätzer $\hat{\mu}_i$. Die (unskalierte) Devianz eines GLMs ist definiert als

$$Dev_l(\hat{\mu} | \mathbf{y}) := -2\varphi[\ell(\hat{\mu} | \mathbf{y}, \varphi) - \ell(\mathbf{y} | \mathbf{y}, \varphi)] = -2 \sum_{i=1}^n \int_{y_i}^{\hat{\mu}_i} \frac{y_i - t}{V(t)} dt$$

und die skalierte Devianz als

$$Dev_s(\hat{\mu} | \mathbf{y}, \varphi) := Dev_l(\hat{\mu} | \mathbf{y}, \varphi) / \varphi.$$

Dabei ist $\ell(\mathbf{y} | \mathbf{y}, \varphi)$ wie in (2.8) definiert, wobei y_i für μ_i eingesetzt wird.

Die Integralschreibweise ist die für GLMs gebräuchlichere. Sie kann aus der Differenz der Log-Likelihoods gewonnen werden, indem ausgenutzt wird, dass die Beobachtungen in einem GLM aus einer einparametrischen Exponentialfamilie mit Dispersionparameter φ stammen und somit

$$\frac{\partial \ell(\mu_i | y_i, \varphi)}{\partial \mu_i} = \frac{\partial}{\partial \vartheta_i} \left(\frac{y_i \vartheta_i - b(\vartheta_i)}{\varphi} \right) \frac{\partial \vartheta_i}{\partial \mu_i} = \frac{1}{V(\mu_i)} \frac{y_i - \mu_i}{\varphi}$$

gilt. In dieser Schreibweise reicht die Kenntnis des Erwartungswertes μ und der Varianzfunktion $V(\cdot)$ zur Bestimmung der Devianz aus, die Kenntnis der exakten Dichte ist nicht

nötig. Für die skalierte Devianz wird zusätzlich noch die Kenntnis des Dispersionsparameters φ benötigt.

Alternativ kann die wie folgt definierte generalisierte Pearson-Statistik zur Bestimmung der Modellgüte benutzt werden.

Definition 4 (*(skalierte) generalisierte Pearson-Statistik*)

Gegeben seien die unabhängigen Beobachtungen $\mathbf{y} = (y_1, \dots, y_n)$ und die durch ein GLM bestimmten Schätzer $\hat{\mu}_i$. Dann ist die (unskalierte) generalisierte Pearson-Statistik definiert als

$$X^2(\hat{\mu} | \mathbf{y}, \varphi) := \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)}$$

und die skalierte Pearson-Statistik als

$$X_s^2(\hat{\mu} | \mathbf{y}, \varphi) := X^2(\hat{\mu} | \mathbf{y}) / \varphi.$$

Ist der Dispersionsparameter φ unbekannt, so kann z.B. der Schätzer aus (2.10) eingesetzt werden.

Beide Gütemaße nehmen nur Werte größer Null an und sind umso kleiner, je größer die Güte des Modells ist. Unter Normalverteilung entsprechen beide Gütemaße der Fehlerquadratsumme. Entstammen die Beobachtungen einer Poisson- oder einer Binomialverteilung, so entspricht die generalisierte Pearson-Statistik der Originalversion $X^2 = \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{\widehat{Var}(y_i)}$. Für die Poisson- und die Gammaverteilung nimmt der Dispersionsparameter nämlich den Wert Eins an und somit ist $\widehat{Var}(y_i) = V(\hat{\mu}_i)$.

Zwar ist die skalierte Devianz als auch die skalierte generalisierte Pearson Statistik unter Normalverteilung χ^2 -verteilt, laut Firth (1991) ist die Devianz im Allgemeinen jedoch weder χ^2 -verteilt noch unabhängig von $\hat{\alpha}$. Für zwei verschachtelte Modelle M_A und M_B der Dimensionen $A < B$ folgt die Differenz der zugehörigen Differenzen jedoch asymptotisch einer χ^2_{B-A} -Verteilung:

$$Dev_s(\hat{\mu}_A | \mathbf{y}) - Dev_s(\hat{\mu}_B | \mathbf{y}) \stackrel{\text{asympt.}}{\sim} \chi^2_{B-A}.$$

Somit kann basierend auf dieser Teststatistik eine schrittweise Variablenselektion für den linearen Prädiktor η durchgeführt werden.

Genau wie die Gütemaße der einfachen linearen Regression nicht für GLMs geeignet sind, müssen auch die Residuen entsprechend den geänderten Voraussetzungen angepasst werden. Die einfachen Residuen $y_i - \hat{\mu}_i$ sind nämlich, wenn die Verteilung der y_i keiner Normalverteilung mehr folgt, nicht mehr geeignet, die Anpassung eines Modells zu überprüfen. Somit werden generalisierte Residuen benötigt, die im Folgenden definiert werden:

Definition 5 (*generalisierte Residuen*)

Gegeben seien die einer einparametrischen Exponentialfamilie mit Dispersionsparameter φ entstammenden, unabhängigen Beobachtungen $\mathbf{y} = (y_1, \dots, y_n)$ und die durch ein GLM bestimmten Schätzer $\hat{\mu}_i$. Dann gibt es die zwei folgenden Arten von generalisierten Residuen:

Pearson-Residuen r_P :

Die Pearson-Residuen sind definiert als Residuen dividiert durch die Wurzel aus der Varianzfunktion an der Stelle $\hat{\mu}_i$:

$$r_P(y_i, \hat{\mu}_i) := \frac{y_i - \hat{\mu}_i}{\sqrt{V(\hat{\mu}_i)}}.$$

Devianz-Residuen r_D :

Die Devianz-Residuen sind definiert durch:

$$r_D(y_i, \hat{\mu}_i) = \text{sign}(y_i - \hat{\mu}_i) \sqrt{-2 \int_{y_i}^{\hat{\mu}_i} \frac{y_i - t}{V(t)} dt}.$$

Der Name der Pearson-Residuen kommt von der Pearson-Statistik, die sich aus der Summe der quadrierten Pearson-Residuen ergibt: $X^2 = \sum_{i=1}^n r_P^2(y_i, \hat{\mu}_i)$. Die Devianz-Residuen sind so definiert, dass ihre aufsummierten quadrierten Werte das zugehörige Gütemaß Devianz ergeben. Auch hier kann φ wieder durch einen Schätzer ersetzt werden.

2.2.3 Quasi-Likelihood

Wird nicht nur ein beliebiger Dispersionsparameter φ unterstellt, sondern sollen auch beliebige Varianzfunktionen $V(\mu_i)$ modelliert werden, muss die Log-Likelihood erweitert werden. Dabei sollen für die Asymptotik wichtige Eigenschaften der Log-Likelihood auch für die Erweiterung gültig sein, siehe Firth (1991). In der Regel beruhen asymptotische

Aussagen auf den folgenden zwei Eigenschaften der Ableitung der Log-Likelihood:

$$\begin{aligned}
 \text{(E1)} \quad E \left(\frac{\partial \ell(\mu_i | y_i, \varphi)}{\partial \mu_i} \right) &= E \left(\frac{y_i - \mu_i}{\varphi V(\mu_i)} \right) = 0 \\
 \text{(E2)} \quad E \left(-\frac{\partial^2 \ell(\mu_i | y_i, \varphi)}{\partial \mu_i^2} \right) &= \text{Var} \left(\frac{\partial \ell(\mu_i | y_i, \varphi)}{\partial \mu_i} \right) \\
 \text{denn } \text{Var} \left(\frac{\partial \ell(\mu_i | y_i, \varphi)}{\partial \mu_i} \right) &= \text{Var} \left(\frac{y_i - \mu_i}{\varphi V(\mu_i)} \right) = \frac{\text{Var}(y_i)}{\varphi^2 V(\mu_i)^2} = \frac{1}{\varphi V(\mu_i)} \\
 \text{und } E \left(-\frac{\partial^2 \ell(\mu_i | y_i, \varphi)}{\partial \mu_i^2} \right) &= E \left[\frac{V(\mu_i) + (y_i - \mu_i) \frac{\partial}{\partial \mu_i} V(\mu_i)}{\varphi V^2(\mu_i)} \right] = \frac{1}{\varphi V(\mu_i)}.
 \end{aligned}$$

Diese Eigenschaften von $\frac{\partial \ell(\mu_i | y_i, \varphi)}{\partial \mu}$ gelten nicht nur unter der Annahme, dass y_i zu einer Exponentialfamilie gehört sondern auch, wenn nur die zwei folgenden Annahmen an die ersten zwei Momente getroffen werden, vorausgesetzt, dass weiterhin $\frac{\partial \ell(\mu_i | y_i, \varphi)}{\partial \mu} = \frac{y_i - \mu_i}{\varphi V(\mu_i)}$ gilt:

$$\text{i) } E(y_i) = \mu_i(\alpha) \quad \text{und} \quad \text{ii) } \text{Var}(y_i) = \varphi V(\mu_i).$$

Somit ist die Quasi-Likelihood QL (genauer wäre eigentlich Quasi-Log-Likelihood) über ihre Ableitung definiert, siehe McCullagh und Nelder (1989), (S. 325):

Definition 6 (*Quasi-Likelihood*)

Für gegebene Einstellungen \mathbf{x}_i der Lokationseffekte seien die unabhängig verteilten Zufallsvariablen y_1, \dots, y_n mit Erwartungswert $\mu_i = g^{-1}(\mathbf{x}_i^T \alpha)$ und Varianz $\varphi V(\mu_i)$ gegeben.

Dann ist die QL $Q(\mu | \mathbf{y}) := Q(\mu | \mathbf{y}, \mathbf{X}, \varphi)$ aller Beobachtungen definiert als

$$Q(\mu | \mathbf{y}) = \sum_{i=1}^n Q(\mu_i | y_i) = \sum_{i=1}^n \left[\int_{y_i}^{\mu_i} \frac{y_i - t}{\varphi V(t)} dt + c(y_i) \right],$$

wobei $c(y_i)$ eine Funktion ist, die weder von μ_i noch von φ abhängt.

Existiert eine Exponentialfamilie, deren Erwartungswert und Varianz genau der Spezifizierung der Quasi-Likelihood entsprechen, so stimmen auch die aus der QL gewonnenen Schätzer mit den echten Log-Likelihood-Schätzern überein. Wenn der Dispersionsparameter φ unbekannt ist, kann er durch einen konsistenten Schätzer, üblicherweise dem Momentenschätzer (2.10), ersetzt werden. Er kann nicht gemeinsam mit dem Parametervektor α aus der QL geschätzt werden, da sich die QL für φ nicht wie eine Log-Likelihood verhält. Es ist anhand von Definition 6 leicht zu sehen, dass $\frac{1}{\varphi}$ vor die Summe gezogen werden kann und sich somit durch Maximierung der QL keine sinnvollen Werte für φ ergeben.

Als Gütemaß wird analog zur Devianz die Quasi-Devianz Dev_Q definiert. Sie eignet sich zum Vergleich von Modellen mit unterschiedlichen Prädiktoren oder auch unterschiedlichen Linkfunktionen, siehe McCullagh und Nelder (1989), S. 327.

$$Dev_Q(\hat{\mu} | \mathbf{y}) = -2\varphi \{Q(\hat{\mu} | \mathbf{y}) - Q(\mathbf{y} | \mathbf{y})\} = -2 \sum_{i=1}^n \int_{y_i}^{\hat{\mu}_i} \frac{y_i - t}{V(t)} dt = \sum_{i=1}^n r_D^2(y_i, \hat{\mu}_i). \quad (2.11)$$

Dies führt zu folgender alternativer Schreibweise der QL:

$$Q(\mu | \mathbf{y}) = -\frac{1}{2} \sum_{i=1}^n \left[\frac{r_D^2(y_i, \mu_i)}{\varphi} + c(y_i) \right]. \quad (2.12)$$

Mit Hilfe der QL kann jetzt eine beliebige Varianzfunktionen $V(\mu)$ und ein beliebiger Dispersionsparameter φ unterstellt werden. Dabei muss die Varianzfunktion jedoch bekannt sein und auch der Dispersionsparameter φ muss entweder bekannt sein oder unabhängig von μ geschätzt werden. Es können jedoch immer noch keine Dispersionsfaktoren unterstellt werden.

2.2.4 Extended-Quasi-Likelihood und Pseudo-Likelihood

Die Quasi-Likelihood muss also um einen Term $q(\cdot)$ erweitert werden, der nur von \mathbf{y} und $\varphi = (\varphi_1, \dots, \varphi_n)^\top$ abhängt. Dieser muss so gestaltet sein, dass die Erweiterung, die so genannte Extended-Quasi-Likelihood (EQL), sich für festes μ wie eine Log-Likelihood für φ verhält. Da der zusätzliche Term unabhängig von μ sein soll, ist gewährt, dass sich die Erweiterung für festes φ weiterhin wie eine QL für μ verhält. In einer ersten Näherung wird angenommen, dass sich die EQL für ein y wie folgt zusammensetzt:

$$\begin{aligned} EQL(\mu, \varphi | y) &= Q(\mu, \varphi | y) + q(\varphi | y) \\ &= -\frac{r_D^2(y, \mu)}{2\varphi} - \frac{1}{2}q_1(\varphi) - q_2(y). \end{aligned}$$

Da die EQL die Eigenschaften einer Log-Likelihood besitzen soll, muss gelten:

$$E\left(\frac{\partial}{\partial \varphi} EQL(\mu, \varphi | y)\right) = \frac{E(r_D^2(\mu, y))}{2\varphi^2} - \frac{1}{2} \frac{\partial}{\partial \varphi} q_1(\varphi) \stackrel{!}{=} 0.$$

Unter Ausnutzung von $E(r_D^2(\mu, y)) \approx \varphi$ folgt, dass $q_1(\varphi) = \log(\varphi) + \text{const.}$ gelten muss.

Die genaue Form der EQL, wie sie in der folgenden Definition angegeben ist, ergibt sich durch eine Sattelpunkt-Approximation der Dichte, siehe McCullagh und Nelder (1989).

Definition 7 (*Extended-Quasi-Likelihood*)

Für die Einstellungen \mathbf{x}_i und \mathbf{z}_i der Lokations- und Dispersionseffekte seien die unabhängig verteilten Zufallsvariablen y_i , $i = 1, \dots, n$, mit Erwartungswert μ_i und Varianz $\varphi_i V(\mu_i)$ gegeben. Weiter sei $\mu_i := \mu_i(\alpha) = g^{-1}(\mathbf{x}_i^\top \alpha)$ und $\varphi_i := \varphi_i(\gamma) = h^{-1}(\mathbf{z}_i^\top \gamma)$ für alle i , $i = 1, \dots, n$, wobei $g : \mathbb{R} \rightarrow \mathbb{R}$ und $h : \mathbb{R} \rightarrow \mathbb{R}$ bijektive beliebig oft differenzierbare

Funktionen sind. Die Extended-Quasi-Likelihood $EQL(\mu, \varphi | \mathbf{y}) := EQL(\mu, \varphi | \mathbf{y}, \mathbf{X}, \mathbf{Z})$ für die Beobachtungen $\mathbf{y} = (y_1, \dots, y_n)^\top$ ist dann definiert als

$$EQL(\mu, \varphi | \mathbf{y}) = -\frac{1}{2} \sum_{i=1}^n \log(2\pi\varphi_i V(y_i)) - \frac{1}{2} \sum_{i=1}^n \frac{r_D^2(y_i, \mu_i)}{\varphi_i}.$$

Die folgenden Überlegungen zeigen, dass die EQL für festes μ die Form einer Log-Likelihood einer Gammaverteilung besitzt:

Die Dichte $f_\Gamma(u_i | a, d_i)$ einer gammaverteilten Zufallsvariablen u_i mit Parametern a und d_i lässt sich wie folgt in Form einer einparametrischen Exponentialfamilie mit natürlichem Parameter ϑ darstellen:

$$\begin{aligned} f_\Gamma(u_i | a, d_i) &= \frac{1}{\Gamma(a)} (ad_i)^a u_i^{a-1} \exp\{-ad_i u_i\} \\ &= \frac{a^a u_i^{a-1}}{\Gamma(a)} \exp\{a(-d_i u_i + \log d_i)\} \\ &\stackrel{\vartheta_i := -d_i}{=} \frac{a^a u_i^{a-1}}{\Gamma(a)} \exp\{a(\vartheta_i u_i + \log(-\vartheta_i))\}. \end{aligned}$$

Somit ergibt sich der Dispersionsparameter τ zu $\frac{1}{a}$. Für $a = \frac{1}{2}$ bzw. $\tau = 2$ ergibt sich die zugehörige Log-Likelihood $\ell(\vartheta_i | u_i)$ zu

$$\ell(\vartheta_i | u_i) = c(u_i) + \frac{1}{2} \vartheta_i u_i + \frac{1}{2} \log(-\vartheta_i).$$

Mit $b(\vartheta_i) = \log(-\vartheta_i)$ ergibt sich für den Erwartungswert $E(u_i) = \frac{\partial}{\partial \vartheta_i} b_i(\vartheta_i) = -\frac{1}{\vartheta_i}$ und für die Varianzfunktion $V(\mu_i) = \frac{\partial^2}{\partial \vartheta_i^2} b_i(\vartheta_i) = \frac{1}{\vartheta_i^2}$. Setzen wir nun die quadrierten Devianzresiduen $r_D^2(y_i, \mu_i)$ für die Beobachtungen u_i ein und setzen $\varphi_i := -\frac{1}{\vartheta_i}$, da φ_i ungefähr dem Erwartungswert von $r_D^2(y_i, \mu_i)$ entspricht, so ergibt sich:

$$\ell(\varphi_i | r_D^2(y_i, \mu_i)) = c(r_D^2(y_i, \mu_i)) - \frac{1}{2} \frac{r_D^2(y_i, \mu_i)}{\varphi_i} - \frac{1}{2} \log(\varphi_i).$$

Somit entspricht die Log-Likelihood eines Gamma GLMs mit log-Link und Dispersionsparameter $\tau = 2$ bis auf eine Konstante der EQL.

Bei Verwendung der EQL wird häufig auch von einem doppelt Generalisierten Linearen Modell (dGLM) gesprochen, da die EQL aufgrund ihrer Konstruktion ein Lokations- und ein Dispersionsmodell miteinander verknüpft. Dabei basiert das Lokationsmodell bei festem φ auf den Beobachtungen $y_i, i = 1, \dots, n$, und ist darstellbar als:

$$\begin{aligned} E(y_i) &= \mu_i, & g(\mu_i) = \eta_i &= \alpha_0 + \sum_{p \in L} \alpha_p x_{ip} \\ \text{Var}(y_i) &= \varphi_i V(\mu_i), & \varphi_i &> 0. \end{aligned} \tag{2.13}$$

Das Dispersionsmodell für festes μ_i beruht, wie schon oben erwähnt, auf den Devianz-Residuen $r_D^2(y_i, \mu_i) =: r_i$ des Lokationsmodells und ist gegeben durch

$$\begin{aligned} E(r_i) &= \varphi_i, & h(\varphi_i) = \xi_i &= \gamma_0 + \sum_{q \in D} \gamma_q z_{iq} \\ \text{Var}(r_i) &= 2\varphi_i^2. \end{aligned} \tag{2.14}$$

In der Regel wird als Linkfunktion $h(\cdot)$ der natürliche Logarithmus gewählt, da das durch ihn implizierte multiplikative Varianzmodell leicht zu interpretieren ist und positive Vorhersagen garantiert.

Das Lokationsmodell entspricht nur dann einem klassischen GLM, wenn es eine Exponentialfamilie gibt, deren Abhängigkeit zwischen Erwartungswert und Varianz obigem Modell

entspricht und wenn $\varphi_i \equiv \varphi$. Dagegen ist das durch die EQL implizierte Dispersionsmodell bei festem μ , wie oben gesehen, ein klassisches GLM mit Gamma-Verteilung und log-Link.

Die Schätzer für α und γ können wieder mit Hilfe des Fisher-Scoring berechnet werden.

Die zugehörigen Scorefunktionen lauten wie folgt:

$$\begin{aligned} \mathcal{S}_{EQL}(\alpha_p) &= \sum_{i=1}^n \frac{\partial EQL(\mu_i, \phi_i | y_i)}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \alpha_p} = \sum_{i=1}^n \frac{y_i - \mu_i}{\phi_i V(\mu_i)} \frac{\partial g^{-1}(\eta_i)}{\partial \eta_i} x_{ip} = 0, \\ \mathcal{S}_{EQL}(\gamma_q) &= \sum_{i=1}^n \frac{\partial EQL(\mu_i, \phi_i | y_i)}{\partial \phi_i} \frac{\partial \phi_i}{\partial \xi_i} \frac{\partial \xi_i}{\partial \gamma_q} = \sum_{i=1}^n \frac{r_D^2(y_i, \mu_i) - \phi_i}{2\phi_i^2} \frac{\partial h^{-1}(\xi_i)}{\partial \xi_i} z_{iq} = 0. \end{aligned}$$

Da, wie man an obigen Scorefunktionen sehen kann, $E\{\partial^2 Q^+ / \partial \alpha_p \partial \gamma_q\} = 0$, $\forall p \in \mathcal{L}$, $q \in \mathcal{D}$ gilt, kann statt einem gemeinsamen Fisher-Scoring auch eine abwechselnde Erneuerung durchgeführt werden, bei der jeweils der aktuelle Schätzer für α bzw. γ festgehalten wird.

Sowohl Nelder und Lee (1991) als auch Aitken (1987) bemerken, dass der Fisher-Scoring-Algorithmus äquivalent als Iteratively Weighted Least Squares (IWLS) dargestellt werden kann. Dabei ergibt sich $\alpha^{(k+1)}$ als gewichteter KQ-Schätzer mit $(\exp(\mathbf{z}_i^T \gamma^{(k)})) V(\mu_i(\alpha^{(k)}))^{-1}$ als Gewichten und $\gamma^{(k+1)}$ ist der Schätzer eines Gamma GLMs mit log-Link und den quadrierten Devianzresiduen $r_d^2(y_i, \mu_i(\alpha^{(k)}))$ als Zielgrößen.

Eine geeignete Wahl der Startwerte für α und φ wird z.B. in Smyth (1989) diskutiert. Demnach werden für α als Startwerte die Koeffizientenschätzer eines einfachen linearen Modells genommen, d.h. es wird $\varphi_i = 1$ gesetzt. Somit ist

$$\alpha_0 = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T g(\mathbf{y}).$$

Als Startwerte für φ werden die Koeffizientenschätzer eines linearen Modells verwendet, bei dem die logarithmierten Residuen des vorher aufgestellten Lokationsmodells als Zielgrößen dienen. Um konsistente Startwerte zu erhalten, wird den logarithmierten Residuen noch 1.27036 hinzuaddiert, so dass sich

$$\varphi_0 = (\mathbf{Z}^\top \mathbf{Z})^{-1} \mathbf{Z}^\top (\log(\mathbf{y} - \hat{\mathbf{y}}) + 1.27036)$$

ergibt, wobei $\hat{\mathbf{y}}$ die Schätzer für \mathbf{y} aus dem Lokationsmodell sind. Smyth et al. (2001) erwähnen, dass Konvergenz schneller erreicht wird, wenn jeweils immer nur ein Schritt des Fisher-Scoring durchgeführt wird, also insbesondere im Gamma GLM.

Der durch die Maximierung der EQL erhaltene Schätzer $\hat{\mu}$ entspricht dem Maximum-Quasi-Likelihood-Schätzer. Bei festem Dispersionsparameter entspricht der Schätzer für φ der mittleren Devianz $\frac{Dev(\hat{\mu}|\mathbf{y})}{n}$ (siehe Nelder und Pregibon (1987)). Im Fall von normalverteilten Fehlern entspricht dies dem ML-Schätzer für φ .

Eine Alternative zur EQL ist die sowohl von Engel und Huele (1996) als auch von Nelder (2000) vorgeschlagene Pseudo-Likelihood (PL). Sie kann als Dichte einer Normalverteilung interpretiert werden, bei der die Varianz wie folgt vom Erwartungswert abhängt: $Var(y_i) = \varphi_i V(\mu_i)$. Alternativ kann die PL auch als EQL aufgefasst werden, bei der jedoch die Varianzfunktion mit in den Dispersionsparameter eingeht, so dass dieser von μ abhängt und als neue Varianzfunktion $V^*(\mu_i) \equiv 1$ resultiert, vgl. Firth (1991). Auf Grund der möglichen Normalverteilungsinterpretation basiert das Dispersionsmodell bei Anwendung der Pseudo-Likelihood immer auf den Pearson-Beiträgen.

Definition 8 (*Pseudo-Likelihood*)

Für die Einstellungen \mathbf{x}_i und \mathbf{z}_i der Lokations- und Dispersionseffekte seien die unabhängig verteilten Zufallsvariablen y_i , $i = 1, \dots, n$, mit Erwartungswert μ_i und Varianz $\varphi_i V(\mu_i)$ gegeben. Weiter sei $\mu_i := \mu_i(\alpha) = g^{-1}(\mathbf{x}_i^\top \alpha)$ und $\varphi_i := \varphi_i(\gamma) = h^{-1}(\mathbf{z}_i^\top \gamma)$ für alle i , $i = 1, \dots, n$, wobei $g : \mathbb{R} \rightarrow \mathbb{R}$ und $h : \mathbb{R} \rightarrow \mathbb{R}$ bijektive beliebig oft differenzierbare Funktionen sind. Die Pseudo-Likelihood $PL(\mu, \varphi | \mathbf{y}) := PL(\mu, \varphi | \mathbf{y}, \mathbf{X}, \mathbf{Z})$ für die Beobachtungen $\mathbf{y} = (y_1, \dots, y_n)^\top$ ist dann definiert als

$$PL(\mu, \varphi | \mathbf{y}) = -\frac{1}{2} \sum_{i=1}^n \log(2\pi\varphi_i V(\mu_i)) - \frac{1}{2} \sum_{i=1}^n \frac{r_P^2(y_i, \mu_i)}{\varphi_i}.$$

Sie wird abkürzend auch mit PL bezeichnet.

Die Scorefunktionen der PL für α und γ lauten

$$\begin{aligned} \mathcal{S}_{PL}(\alpha_p) &:= \frac{\partial}{\partial \alpha_p} PL(\mu, \varphi | \mathbf{y}) = \sum_{i=1}^n \frac{\partial PL(\mu_i, \varphi_i | y_i)}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \alpha_p} \\ &= \frac{1}{2} \sum_{i=1}^n \frac{1}{\varphi_i V(\mu_i)} \left[(r_P^2(y_i, \mu_i) - \varphi_i) \frac{\partial V(\mu_i)}{\partial \mu_i} + 2(y_i - \mu_i) \right] \frac{\partial g^{-1}(\eta_i)}{\partial \eta_i} x_{ip}, \\ \mathcal{S}_{PL}(\gamma_q) &:= \frac{\partial}{\partial \gamma_q} PL(\mu, \varphi | \mathbf{y}) = \sum_{i=1}^n \frac{\partial PL(\mu_i, \varphi_i | y_i)}{\partial \varphi_i} \frac{\partial \varphi_i}{\partial \xi_i} \frac{\partial \xi_i}{\partial \gamma_q} \\ &= \sum_{i=1}^n \frac{r_P^2(y_i, \mu_i) - \varphi_i}{2\varphi_i^2} \frac{\partial h^{-1}(\xi_i)}{\partial \xi_i} z_{iq}. \end{aligned}$$

Die für die EQL festgestellte Orthogonalität überträgt sich deshalb bei $V(\mu_i) \neq 1$ nicht auf die PL, sondern wir erhalten für $p \in \mathcal{L}$ und $q \in \mathcal{D}$:

$$E \left\{ -\frac{\partial^2 PL(\mu, \varphi | \mathbf{y})}{\partial \alpha_p \partial \gamma_q} \right\} = \frac{1}{2} \sum_{i=1}^n \frac{1}{\varphi_i V(\mu_i)} \frac{\partial V(\mu_i)}{\partial \mu_i} \frac{\partial g^{-1}(\eta_i)}{\partial \eta_i} \frac{\partial h^{-1}(\xi_i)}{\partial \xi_i} x_{ip} z_{iq}.$$

Somit ist hier ein gemeinsames Fisher-Scoring statt einer abwechselnden Erneuerung nötig. Dabei werden für α und γ die gleichen Startwerte wie bei Maximierung der EQL verwendet.

Um mit Hilfe der EQL oder PL gewonnene Modelle mit unterschiedlichen Einflussfaktoren für die Dispersion und Lokation miteinander zu vergleichen, ist es wünschenswert, das zuvor definierte Gütemaß Devianz zu verallgemeinern. Sowohl im Log-Likelihood- als auch im Quasi-Likelihood-Ansatz werden nur die Erwartungswerte μ_i modelliert. Die Devianz ermöglicht es, analog zum Bestimmtheitsmaß in der linearen Regression, einen Vergleich der vorhergesagten Erwartungswerte $\hat{\mu}_i$ mit den beobachteten Werten y_i durchzuführen. Sowohl im EQL- als auch im PL-Ansatz werden jedoch auch die Varianzen $\varphi_i V(\mu_i)$ modelliert. Somit wäre der Vergleich zwischen vorhergesagten Erwartungswerten und Beobachtungen auf einen Vergleich zwischen den vorhergesagten Varianzen und den empirischen Varianzen zu erweitern. Hierzu gibt es jedoch keine Beobachtungen, so dass bei dieser Art der Modellierung ein sinnvoller Bezugswert fehlt. Daher wird die Extended-Quasi-Devianz definiert als $Dev_{EQL}(\hat{\mu}, \hat{\varphi} | \mathbf{y}) := -2 EQL(\hat{\mu}, \hat{\varphi} | \mathbf{y})$ und die Pseudo-Devianz als $Dev_{PL}(\hat{\mu}, \hat{\varphi} | \mathbf{y}) := -2 PL(\hat{\mu}, \hat{\varphi} | \mathbf{y})$, vgl. McCullagh und Nelder (1989), Abschnitte 10.4 und Nelder und Lee (1991). Durch den fehlenden Bezugswert können die Devianzen im EQL- und PL-Ansatz auch negativ werden und sind somit isoliert betrachtet nur bedingt interpretierbar.

2.2.5 Restringierte Versionen der Extended Quasi-Likelihood und der Pseudo-Likelihood

Da in nichtwiederholten Plänen keine Beobachtungen für die Varianz vorliegen, erfolgt die Varianzschätzung auf Basis der Residuen des Lokationsmodells. Diese Anpassung führt jedoch, in Abhängigkeit der Anzahl der geschätzten Parameter, zu einer Reduzierung der durchschnittlichen Residuengröße und somit zu einer Unterschätzung der wahren Varianz. Somit muss eine Adjustierung für die Freiheitsgrade des Lokationsmodells vorgenommen werden, um verbesserte Schätzer für die Dispersionseffekte zu erhalten.

Eine Möglichkeit, restringierte Schätzer zu erhalten, stammt von McCullagh und Nelder (1989) S. 363. Sie versehen den vorderen Teil der EQL mit dem Adjustierungsfaktor $\frac{n-L-1}{n}$, um die Anzahl L der ins Lokationsmodell eingehenden Effekte und das allgemeine Mittel zu berücksichtigen. Laut McCullagh und Nelder (1989) S. 363 führt dieses Vorgehen approximativ zu restringierten ML-Schätzern. Für ein Modell mit konstantem φ führt dies zu einem Schätzer $\hat{\varphi}$, der im Vergleich zur unrestringierten Version mit $\frac{n}{n-L-1}$ multipliziert wird, also eine zu gute Anpassung des Lokationsmodells ausgleicht. Dies entspricht dem unverzerrten Schätzer $\hat{\varphi}$, siehe 2.10. Unter der üblichen Annahme, dass die Dispersionslinkfunktion der Logarithmus ist, führt die Nutzung dieser restringierten EQL-Version nur zu veränderten Schätzern für das allgemeine Mittel des Dispersionsmodells, die restlichen Koeffizienten des Dispersionsmodells bleiben unverändert.

Dieses Vorgehen kann auf die PL übertragen werden, indem der Adjustierungsfaktor an die entsprechende Stelle der PL gesetzt wird. Somit resultieren folgende restringierte

Versionen der *EQL* und der *PL*:

Definition 9 (*restringierte EQL- und PL-Versionen nach McCullagh und Nelder*)

Für die Einstellungen \mathbf{x}_i und \mathbf{z}_i der Lokations- und Dispersionseffekte seien die unabhängig verteilten Zufallsvariablen $y_i, i = 1, \dots, n$, mit Erwartungswert μ_i und Varianz $\varphi_i V(\mu_i)$ gegeben. Weiter sei $\mu_i := \mu_i(\alpha) = g^{-1}(\mathbf{x}_i^\top \alpha)$ und $\varphi_i := \varphi_i(\gamma) = h^{-1}(\mathbf{z}_i^\top \gamma)$ für alle $i, i = 1, \dots, n$, wobei $g : \mathbb{R} \rightarrow \mathbb{R}$ und $h : \mathbb{R} \rightarrow \mathbb{R}$ bijektive beliebig oft differenzierbare Funktionen sind. Die restringierte Extended Quasi-Likelihood $EQL^M(\mu, \varphi | \mathbf{y}) := EQL^M(\mu, \varphi | \mathbf{y}, \mathbf{X}, \mathbf{Z})$ für alle Beobachtungen von McCullagh und Nelder lautet dann:

$$EQL^M(\mu, \varphi | \mathbf{y}) = -\frac{n-L-1}{2n} \sum_{i=1}^n \log(2\pi\varphi_i V(y_i)) - \frac{1}{2} \sum_{i=1}^n \frac{r_D^2(y_i, \mu_i)}{\varphi_i}.$$

Die restringierte Pseudo-Likelihood $PL^M(\mu, \varphi | \mathbf{y}) := PL^M(\mu, \varphi | \mathbf{y}, \mathbf{X}, \mathbf{Z})$ für alle Beobachtungen nach McCullagh und Nelder ist definiert als

$$PL^M(\mu, \varphi | \mathbf{y}) = -\frac{n-L-1}{2n} \sum_{i=1}^n \log(2\pi\varphi_i V(\mu_i)) - \frac{1}{2} \sum_{i=1}^n \frac{r_P^2(y_i, \mu_i)}{\varphi_i}.$$

Sie werden abkürzend auch mit EQL^M bzw. PL^M bezeichnet.

Ein weiterer Vorschlag zum Erhalt von restringierten Schätzern stammt von Lee und Nelder (1998). Sie schlagen in Anlehnung an Cox und Reid (1987) folgende *EQL*-Variante vor:

$$\begin{aligned} EQL^C(\mu, \varphi | \mathbf{y}, \mathbf{X}, \mathbf{Z}) &:= EQL^C(\mu, \varphi | \mathbf{y}) \\ &= -\frac{1}{2} \sum_{i=1}^n \log(2\pi\varphi_i V(y_i)) - \frac{1}{2} \sum_{i=1}^n \frac{r_D^2(y_i, \mu_i)}{\varphi_i} \\ &\quad - \log \left\{ \det \left[\frac{\mathbf{X}^\top \text{diag} \left(\frac{1}{\varphi_i V(\mu_i)} \frac{\partial \mu_i}{\partial \eta_i} \right) \mathbf{X}}{2\pi} \right] \right\}. \end{aligned} \tag{2.15}$$

Da die Maximierung dieser Funktion Computer-intensiv ist und nur langsam konvergiert, schlagen Lee und Nelder (1998) vor, die Dispersionsparameter in den ersten $k-1$ Schritten basierend auf der Approximation

$$\begin{aligned} EQL^L(\mu, \varphi | \mathbf{y}, \mathbf{X}, \mathbf{Z}) &:= EQL^L(\mu, \varphi | \mathbf{y}) \\ &= -\frac{1}{2} \sum_{i=1}^n \log(2\pi\varphi_i V(y_i)) - \frac{1}{2} \sum_{i=1}^n \frac{r_D^2(y_i, \mu_i)}{\varphi_i(1-h_{ii})} \end{aligned} \quad (2.16)$$

zu berechnen, wobei $h_{ii} = \left(W^{1/2} D X (X^T D W D X)^{-1} X^T D W^{1/2} \right)_{ii}$ mit

$$W = \text{diag} \left(\frac{1}{\varphi_i V(\mu_i)} \right) \quad \text{und} \quad D = \text{diag} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)$$

und nur im letzten Schritt die EQL^C zu maximieren. Die Schätzer für α werden dabei jedoch weiterhin auf Basis der unrestringierten EQL bestimmt. Die Maximierung der EQL^L entspricht der Anpassung eines Gamma-Dispersions-Modells mit Zielgrößen $r_D^2(y_i, \mu_i)/(1-h_{ii})$ anstelle von $r_D^2(y_i, \mu_i)$.

Im Folgenden wird die restringierte Pseudo-Likelihood nach Lee und Nelder PL^L analog zur restringierten EQL-Version von Lee und Nelder (1998) definiert:

Definition 10 (*restringierte Pseudo-Likelihood nach Lee und Nelder*)

Für die Einstellungen \mathbf{x}_i und \mathbf{z}_i der Lokations- und Dispersionseffekte seien die unabhängig verteilten Zufallsvariablen $y_i, i = 1, \dots, n$, mit Erwartungswert μ_i und Varianz $\varphi_i V(\mu_i)$ gegeben. Weiter sei $\mu_i := \mu_i(\alpha) = g^{-1}(\mathbf{x}_i^T \alpha)$ und $\varphi_i := \varphi_i(\gamma) = h^{-1}(\mathbf{z}_i^T \gamma)$ für alle $i, i = 1, \dots, n$, wobei $g : \mathbb{R} \rightarrow \mathbb{R}$ und $h : \mathbb{R} \rightarrow \mathbb{R}$ bijektive beliebig oft differenzierbare Funktionen sind. Die restringierte Pseudo-Likelihood nach Lee und Nelder

$PL^L(\mu, \varphi | \mathbf{y}) := PL^L(\mu, \varphi | \mathbf{y}, \mathbf{X}, \mathbf{Z})$ für alle Beobachtungen ist definiert als

$$PL^L(\mu, \varphi | \mathbf{y}) = -\frac{1}{2} \sum_{i=1}^n \log(2\pi\varphi_i V(\mu_i)) - \frac{1}{2} \sum_{i=1}^n \frac{r_P^2(y_i, \mu_i)}{\varphi_i(1 - h_{ii})}$$

mit $h_{ii} = (W^{1/2}DX(X^T DW DX)^{-1}X^T DW^{1/2})_{ii}$ wobei

$$W = \text{diag} \left(\frac{1}{\varphi_i V(\mu_i)} \right) \quad \text{und} \quad D = \text{diag} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)$$

Sie wird abkürzend auch mit PL^L bezeichnet.

Da die PL nicht die Orthogonalitätseigenschaften der EQL besitzt und somit für α und γ ein gemeinsames Fisher-Scoring nötig ist, wird in diesem Fall die restringierte Variante nicht nur zur Schätzung von γ sondern auch von α verwendet. Dabei ändern sich die Schätzer für α im Vergleich zur unrestringierten Variante kaum. Beide restringierte Varianten haben jedoch einen deutlichen Einfluss auf die Schätzung von γ_0 .

Beispielhaft wird folgendes Modell betrachtet:

$$\lambda = 1 - \beta = 2$$

$$E(y_i) = \mu_i = g(10 + A - 0.5B + 0.5A : B)$$

$$\text{Var}(y_i) = \varphi_i V(\mu_i) = \exp(A)\mu_i^{-2}$$

Dabei wird für $g(\cdot)$ die Power-Transformation (2.3) gewählt und den Fehlern eine Normalverteilung unterstellt. Nun wurden für ein faktorielles 2^5 -Design 10 000 Datensätze basierend auf obigen Annahmen simuliert. In Grafik 2.1 sind die zugehörigen Schätzer bei Unterstellung des korrekten Transformationsparameters unter Verwendung der PL und

der restringierten Versionen PL^M und PL^L dargestellt. Basierend auf den restringierten Versionen wird γ_0 deutlich besser geschätzt als bei Verwendung der PL . Wie erwartet, wird γ_0 und somit auch die Varianz von der PL unterschätzt. Für dieses Beispiel ist nur ein geringer Unterschied zwischen der PL^L und der deutlich einfacheren PL^M zu erkennen. Für die anderen Parameter sind keine Änderungen sichtbar, siehe Anhang C.2, Grafiken C.1 bis C.5.

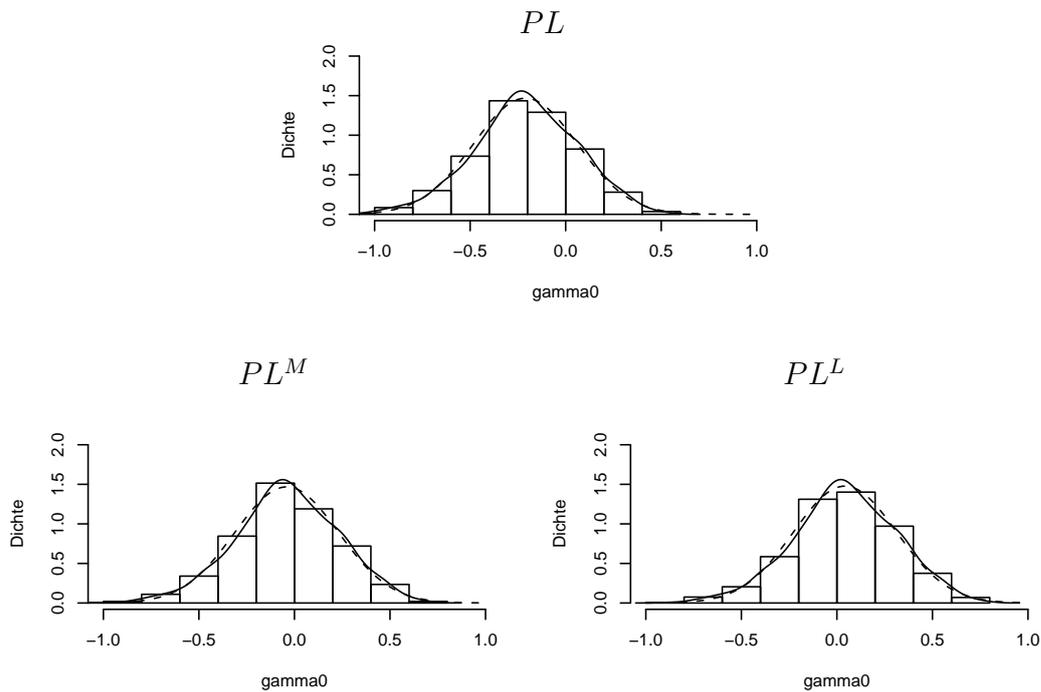


Abbildung 2.1: Histogramm mit Dichteschätzung (—) und Normalverteilungsdichte (- - -) für

γ_0

2.2.6 GLM basierte Methode zur Schätzung von Dispersionseffekten und Transformationsparametern

Mit Hilfe der oben vorgestellten GLM können für nichtwiederholte Pläne sowohl die Koeffizienten der Lokations- als auch der Dispersionseffekte geschätzt werden. Dabei wird angenommen, dass die Varianzfunktion $V(\mu_i)$ und die Linkfunktion $g(\mu_i)$ bekannt sind. In der Praxis ist dies jedoch häufig nicht der Fall. Erdbrügge und Auer (2010) setzen nun folgende Varianz- und Linkfunktion voraus:

$$V(\mu_i) = \mu_i^{2\beta} \quad \text{und} \quad g(\mu_i) = \begin{cases} \frac{\mu_i^\lambda - 1}{\lambda} & : \lambda \neq 0 \\ \log(\mu_i) & : \lambda = 0 \end{cases}, \quad (2.17)$$

wobei λ und β unbekannt und unabhängig sind.

Somit ergibt sich die PL zu

$$PL(\alpha, \gamma, \lambda, \beta) = -\frac{1}{2} \sum_{i=1}^n \left(\log(2\pi) + \mathbf{z}_i^\top \gamma + 2\beta \log \mu_i + \frac{(y_i - \mu_i)^2}{\exp(\mathbf{z}_i^\top \gamma) \mu_i^{2\beta}} \right), \quad \text{mit} \quad (2.18)$$

$$\mu_i = g^{-1}(\mathbf{x}_i^\top \alpha) = \begin{cases} (\mathbf{x}_i^\top \alpha + 1)^{\frac{1}{\lambda}} & : \lambda \neq 0 \\ \exp(\mathbf{x}_i^\top \alpha) & : \lambda = 0 \end{cases}. \quad (2.19)$$

Kann davon ausgegangen werden, dass, wie in den Transformationsmodellen unterstellt, eine einzige Transformation zu additiven Effekten und einer vom Erwartungswert unabhängigen Varianz führen würde, ist $\lambda = 1 - \beta$ zu setzen.

Nun können zusätzlich zu den Koeffizienten der Lokations- und Dispersionsmodelle auch die Parameter λ und β durch die Maximierung einer der Log-Likelihood Varianten geschätzt werden.

Erdbrügge und Auer (2010) zeigen mit Hilfe von Satz 2, dass die durch die Maximierung der PL gewonnenen Schätzer für α, γ, β und λ konsistent und asymptotisch effizient und normalverteilt sind.

Wird unter der Annahme, dass $\lambda = 1 - \beta$, zur Schätzung des Parametervektors $(\alpha, \gamma, \lambda)^T$ der Fisher Scoring Algorithmus verwendet, ist der Schätzer für λ sehr stark von dem gewählten Startwert abhängig. Er weicht meist nur um weniger als eine Nachkommastelle vom beliebig gewählten Startwert ab.

Sinnvoller ist es, für eine Sequenz $\Lambda = \{\lambda_1, \dots, \lambda_r\}$ von λ den Fisher Scoring Algorithmus für α und γ laufen zu lassen. Der Schätzvektor ergibt sich dann aus dem zur kleinsten Devianz gehörenden λ -Wert und den zugehörigen Schätzwerten für α und γ .

Ähnliches ist zu erwarten, wenn β unabhängig von λ geschätzt werden soll. Somit sollte in diesem Fall der Fisher-Scoring Algorithmus für Sequenzen $\Lambda = \{\lambda_1, \dots, \lambda_r\}$ von λ und $\Gamma = \{\gamma_1, \dots, \gamma_l\}$ von γ durchgeführt werden.

Alternativ kann λ auch ähnlich zum grafischen Verfahren von Box und Cox mit Hilfe einer Grafik geschätzt werden. Anstatt λ gegen die SSE abzutragen, wird λ gegen die Devianz abgetragen. Beides sind Gütekriterien die - in den jeweiligen Modellvorstellungen - äquivalent zur Maximierung der Log-Likelihood minimiert werden können. Erdbrügge und Auer (2010) schlagen vor, gleichzeitig diese Grafik zur Identifizierung von aktiven Dispersionsfaktoren zu nutzen. Wird unterstellt, dass es maximal nur einen aktiven Dispersionsfaktor gibt, wird dabei wie folgt vorgegangen: Zuerst werden unter Vernachlässigung eventueller Dispersionseffekte, d.h. mit Standardmethoden wie dem Half-Normal-Plot aktive Lokationseffekte identifiziert. Danach wird die Devianz für alle Dispersionsmodelle mit einem

Haupteffekt und für das Dispersionsmodell mit reinem Intercept gegen λ abgetragen. Weist eines der Modelle mit einem Haupteffekt eine deutliche Verringerung der Devianz im Vergleich zu den anderen Modellen auf, deutet dies darauf hin, dass der zugehörige Effekt aktiv ist. λ wird dann durch den Wert geschätzt, der die Devianz des unterstellten Dispersionsmodells minimiert. Im Folgenden wird eine Grafik, die die Devianz für unterschiedliche Modelle gegen λ abträgt, Devianz Plot genannt.

In Grafik 2.2 ist ein Devianz Plot zu sehen, der auf einen Dispersionseffekt von x_1 und einen Transformationsparameter von $\lambda = 0.25$ hindeutet.

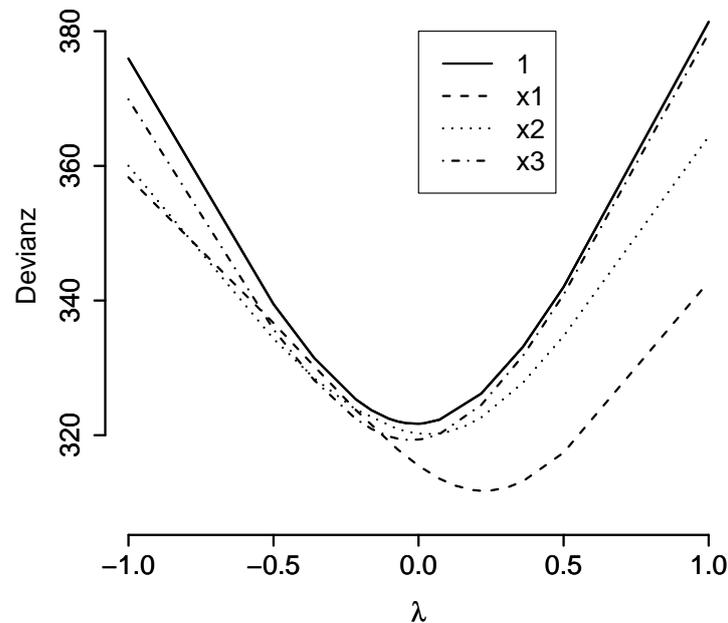


Abbildung 2.2: Devianzplot

Im allgemeinen Fall, dass λ unabhängig von β ist, können λ und β mit Hilfe von einem Konturplot für die Devianz gleichzeitig geschätzt werden. Nelder und Pregibon verwenden diese Konturplots schon zur Überprüfung der Annahme $\lambda = 1 - \beta$ und Erdbrügge (2002)

zur Identifizierung von Dispersionseffekten.

2.3 Erweiterung der Box-Cox-Methode

Kann davon ausgegangen werden, dass mit einer einzigen Transformation die Annahmen der linearen Modelle erfüllt werden können, kann auch die von Erdbrügge und Auer (2010) vorgeschlagene Erweiterung des Box-Cox-Verfahrens, welches auch im Fall von nichtwiederholten Plänen den Transformationsparameter $\lambda = 1 - \beta$ sowie die Parametervektoren der Lokations- und Dispersionseffekte gleichzeitig schätzen kann, verwendet werden. Bei diesem Modell handelt es sich zwar um kein theoretisch korrektes Modell, da der Träger der Verteilung von \mathbf{y} vom Transformationsparameter λ abhängt, siehe Erdbrügge (2003) Abschnitt 2.2.2. Es liefert jedoch approximativ die gleichen Ergebnisse, wie das auf den GLM basierende Verfahren, wenn bei diesem $\lambda = 1 - \beta$ vorausgesetzt wird. Außerdem benötigt es auf Grund des einfacheren Algorithmus deutlich kürzere Rechenzeiten und führt seltener zu Konvergenzproblemen.

Die vorgeschlagene Erweiterung unterstellt das schon in Abschnitt 2.1.2 erwähnte Modell der multiplikativen Heteroskedastizität für die transformierten Daten

$$T_\lambda(\mathbf{y}) = \mathbf{X}\alpha + \mathbf{e}, \quad (2.20)$$

wobei $\mathbf{e} \sim N(\mathbf{0}, \Sigma)$ und $\Sigma = \text{diag}(\exp\{\mathbf{z}_1^T \boldsymbol{\gamma}\}, \dots, \exp\{\mathbf{z}_n^T \boldsymbol{\gamma}\})$,

welches für die Originaldaten folgende Varianz impliziert:

$$\text{Var}(y_i) = \mu_i^{2(1-\lambda)} \exp\{\mathbf{z}_i^T \boldsymbol{\gamma}\}, \quad i = 1, \dots, n.$$

Somit liegt für die Erweiterung der Box-Cox-Methode ein klassisches lineares Modell in den transformierten Daten vor. Die Kovarianzmatrix Σ besitzt jetzt jedoch die Diagonalelemente $\exp\{\mathbf{z}_i^T \boldsymbol{\gamma}\}$ und wird somit direkt durch die Dispersionseffekte beeinflusst. Äquivalent zur Maximierung der gemeinsamen Log-Likelihood kann eine Minimierung der Devianz erfolgen, die für das Modell (2.20) wie folgt definiert ist:

$$\begin{aligned} Dev_{BC}(\lambda, \boldsymbol{\alpha}, \boldsymbol{\gamma}, | \mathbf{y}) &= -2 \ell(\lambda, \boldsymbol{\alpha}, \boldsymbol{\gamma} | \mathbf{y}) \\ &= n \log\{2\pi\} + \sum_{i=1}^n \log \sigma_i^2 + 2(1 - \lambda) \sum_{i=1}^n \log y_i + \sum_{i=1}^n \frac{1}{\sigma_i^2} (T_\lambda(y_i) - \mu_i)^2, \end{aligned}$$

mit $\sigma_i^2 = \exp\{\mathbf{z}_i^T \boldsymbol{\gamma}\}$.

Analog zur Originalmethode kann auch hier eine Vereinfachung erreicht werden, wenn mit der standardisierten Transformation $\tilde{T}_\lambda(\cdot)$ gearbeitet wird. Die resultierende Devianz lautet dann wie folgt:

$$\begin{aligned} \widetilde{Dev}_{BC}(\lambda, \boldsymbol{\alpha}^*, \boldsymbol{\gamma}^*, | \mathbf{y}) &:= -2 \ell(\lambda | \boldsymbol{\alpha}^*, \boldsymbol{\gamma}^*, \mathbf{y}) \\ &= n \log(2\pi) + \sum_{i=1}^n \mathbf{z}_i^T \boldsymbol{\gamma}^* + \sum_{i=1}^n \frac{(\tilde{T}_\lambda(y_i) - \mathbf{x}_i^T \boldsymbol{\alpha}^*)^2}{\exp\{\mathbf{z}_i^T \boldsymbol{\gamma}^*\}}, \end{aligned}$$

wobei

$$\begin{aligned} \boldsymbol{\gamma}^* &= (\gamma_0^*, \gamma_1, \dots, \gamma_q)^T \quad \text{mit} \quad \gamma_0^* = \gamma_0 + 2(1 - \lambda) \log \dot{y} \quad \text{und} \\ \boldsymbol{\alpha}^* &= \boldsymbol{\alpha} \dot{y}^{1-\lambda}. \end{aligned}$$

Erdbrügge und Auer (2010) zeigen, dass die ursprüngliche Methode von Box und Cox in der Erweiterung enthalten ist.

Bei der Anwendung der erweiterten Box-Cox-Methode können die unbekannt Parameter nicht mehr analytisch bestimmt werden. Aber es ist leicht zu sehen, dass die Devianz \widetilde{Dev}_{BC} der EQL entspricht, wenn $\tilde{T}_\lambda(y_i)$ als Zielgrößen verwendet werden und $V(\mu_i) \equiv 1$ gesetzt wird. Somit kann der dort beschriebene IWLS-Algorithmus zur Schätzung von α und γ für festes λ verwendet werden. Des Weiteren kann auch hier die grafische Methode von Box und Cox (1964) erweitert werden, indem die Devianz $\widetilde{Dev}_{BC}(\lambda | \hat{\alpha}, \hat{\gamma}, \mathbf{y})$ gegen eine Sequenz Λ abgetragen wird.

Die EQL wird in der Literatur häufig der PL vorgezogen. Davidian und Carrol (1988) argumentieren zwar wegen ihrer asymptotischen Eigenschaften für die PL. Nelder und Lee (1992) stellen jedoch fest, dass die EQL für kleine Stichproben im Allgemeinen bessere Ergebnisse liefert als die PL. Nelder (2000) sowie Lee und Nelder (2003) lehnen die PL ab, da diese auf der Normalverteilungsannahme beruht und somit Residuen mit einer schiefen Verteilung nicht adäquat modellieren kann. Auch Firth (1991) betont, dass bei Verwendung der PL mit Effizienzverlusten zu rechnen ist, wenn die Beobachtungen einer sehr schiefen Verteilung entstammen. Des Weiteren wurde in Abschnitt 2.2.5 deutlich, dass die Anwendung restringierter Methoden zu besseren Varianzschätzern führt, und somit auch betrachtet werden sollte. Aufgrund dieser Überlegungen soll im Weiteren untersucht werden, ob mit Hilfe von Satz 2 die Konsistenz und asymptotische Effizienz und Normalverteilung der Schätzer auch bei Verwendung der anderen Varianten der Log-Likelihood gelten.

Des Weiteren soll untersucht werden, ob die Idee, Dispersionseffekte grafisch basierend auf der Devianz zu identifizieren, auch auf die Lokationseffekte und in ein formal korrektes Verfahren übertragen werden kann. Dabei sollten auch die Abhängigkeiten zwischen Lokations- und Dispersionsmodell berücksichtigt werden. Auf Grund der schnelleren Rechenzeiten und der geringeren Konvergenzprobleme soll dies auf Basis der erweiterten Box-Cox-Methode geschehen. Wir beschränken uns somit auf den Fall $\lambda = 1 - \beta$.

Kapitel 3

Untersuchungen der asymptotischen Eigenschaften der Log-Likelihood Erweiterungen

In diesem Kapitel sollen die verschiedenen Varianten des auf generalisierten linearen Modellen basierenden Verfahrens auf ihre theoretischen Eigenschaften hin untersucht werden. Erdbrügge und Auer (2010) weisen die Konsistenz, und die asymptotische Normalverteilung und Effizienz der Schätzer für die PL-Version her. Dies ist der einfachste Fall, da der Pseudo-Likelihood eine echte Dichte zu Grunde liegt. Dies ist für die EQL und ihre restringierten Versionen sowie für die restringierten Methoden der PL nicht gegeben. In diesen Fällen schlagen Nelder und Pregibon (1987), Nelder und Lee (1991) sowie Lee und Nelder (2000) vor, für theoretische Untersuchungen eine so genannte Quasi-Verteilung zu benutzen. Dazu wird die interessierende Log-Likelihood-Variante nach Anwendung der

Exponentialfunktion mit einem geeigneten Vorfaktor normalisiert, so dass das Integral Eins ergibt und eine echte Dichte vorliegt. Nelder und Pregibon (1987) sowie Lee und Nelder (2000) weisen darauf hin, dass der Vorfaktor häufig von den zu schätzenden Faktoren abhängt, und somit die Schätzer unter Verwendung der Quasi-Verteilung nicht mehr den Originalschätzern entsprechen. Sie relativieren diese Tatsache jedoch, indem sie bemerken, dass der Vorfaktor meist nur wenig Variabilität in den zu schätzenden Parametern aufweist und somit nur einen geringen Einfluss auf die Schätzwerte hat. Für praktische Fälle empfehlen Nelder und Pregibon (1987) sowie Efron (1986) daher die nichtnormalisierte Form zu verwenden, da diese einfacher anzuwenden ist.

Im Folgenden wird untersucht, inwieweit die theoretischen Eigenschaften der PL-Schätzer auch für die auf der EQL und ihren restringierten Versionen sowie für die auf den restringierten Methoden der PL basierenden Schätzer nachweisbar sind. Dies erfolgt, wie bei den PL-Schätzern in Erdbrügge und Auer (2010), mit Hilfe von Satz 2. Dieser beruht jedoch auf den Eigenschaften der Log-Likelihood, weshalb jeweils zunächst die Quasi-Verteilung zu bestimmen ist. Dann kann überprüft werden, ob die theoretischen Eigenschaften für die aus der Maximierung der zugehörigen Quasi-Log-Likelihood bestimmten Schätzern gegeben sind. Danach wird versucht, Rückschlüsse auf die ursprünglichen interessierenden Schätzer zu ziehen. Dabei werden die in (2.19) unterstellten Funktionen für $g(\cdot)$ und $V(\cdot)$ verwendet.

3.1 Extended-Quasi-Likelihood und ihre restringierten Versionen

3.1.1 Extended-Quasi-Likelihood

Zur Bestimmung der Quasi-Verteilung der EQL ist es somit nötig, das Integral von $\exp(EQL(\mu, \varphi, \lambda, \beta | \mathbf{y}))$ zu bestimmen.

Mit der Varianzfunktion $V(y) = y^{2\beta}$ ergibt sich für die Devianzresiduen folgende Darstellung:

$$\begin{aligned}
 r_D^2(y, \mu) &= -2 \int_y^\mu \frac{y-t}{V(t)} dt \\
 &= -2 \int_y^\mu yt^{-2\beta} - t^{-2\beta+1} \\
 &= -2 \left[\frac{1}{-2\beta+1} yt^{-2\beta+1} - \frac{1}{-2\beta+2} t^{-2\beta+2} \right]_y^\mu \\
 &= -2\mu^{-2\beta} \left(\frac{\mu y}{1-2\beta} - \frac{\mu^2}{2-2\beta} \right) + 2y^{-2\beta} \left(\frac{y^2}{1-2\beta} - \frac{y^2}{2-2\beta} \right) \\
 &= \frac{1}{1-\beta} \mu^{2-2\beta} - \frac{2}{1-2\beta} \mu^{-2\beta+1} y + \left(\frac{2}{1-2\beta} - \frac{1}{1-\beta} \right) y^{2-2\beta}.
 \end{aligned}$$

Hiermit lässt sich die EQL für ein beliebiges i wie folgt schreiben:

$$\begin{aligned}
 EQL(\mu_i, \varphi_i, \lambda, \beta | y_i) &= -\frac{1}{2} \log(2\pi \varphi_i V(y_i)) - \frac{1}{2} \frac{r_D^2(y_i, \mu_i)}{\varphi_i} \\
 &= -\frac{1}{2} \log(2\pi \exp(\mathbf{z}_i^\top \boldsymbol{\gamma}) y_i^{2\beta}) - \frac{1}{\varphi} \left[\frac{1}{2-2\beta} \mu_i^{2-2\beta} - \frac{1}{1-2\beta} \mu_i^{-2\beta+1} y_i \right. \\
 &\quad \left. + \left(\frac{1}{1-2\beta} - \frac{1}{2-2\beta} \right) y_i^{2-2\beta} \right] \exp(\mathbf{z}_i^\top \boldsymbol{\gamma}).
 \end{aligned}$$

Das Integral, welches zur Bestimmung der Quasi-Verteilung gelöst werden muss, lautet also

$$\begin{aligned} & \int_{-\infty}^{\infty} \exp(EQL(\mu_i, \varphi_i, \lambda, \beta | y_i)) dy_i \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} \mathbf{z}_i^T \boldsymbol{\gamma}\right) y_i^{-\beta} \exp\left\{\frac{1}{\varphi} \left[-\frac{1}{2-2\beta} \mu_i^{2-2\beta} + \frac{1}{1-2\beta} \mu_i^{-2\beta+1} y_i\right.\right. \\ & \quad \left.\left. - \left(\frac{1}{2-2\beta} - \frac{1}{1-2\beta}\right) y_i^{2-2\beta}\right]\right\} \exp(\mathbf{z}_i^T \boldsymbol{\gamma}) dy_i. \end{aligned}$$

Um eine Quasi-Verteilung der EQL zu bestimmen, ist somit ein Integral folgender Art zu lösen: $\int ay^{-\beta} \exp(b + cy + dy^{2-2\beta}) dy$.

Eine explizite Angabe der zugehörigen Stammfunktion erscheint somit nicht möglich. Eine numerische Lösung ist zwar im konkreten Fall bestimmbar, trägt aber nicht zur Herleitung von theoretischen Eigenschaften der EQL bei. Somit können theoretische Eigenschaften der EQL-Schätzer zumindest auf diesem Weg nicht untersucht werden.

3.1.2 Restringierte Extended-Quasilikelihood-Versionen

Auch für die restringierten Varianten der EQL sind Integrale obiger Form zu lösen. Für die restringierte Version von McCullagh und Nelder ändert sich nur der Term a : Er wird im Vergleich zur EQL noch mit $\exp\left(\frac{n-L-1}{n}\right)$ multipliziert. Dagegen ändern sich für die restringierte Version nach Lee und Nelder die Terme b, c und d . Sie werden jeweils noch durch $(1 - h_{ii})$ dividiert. Somit ist es auch in diesen Fällen nicht möglich, eine Quasi-Verteilung zu bestimmen.

3.2 Restringierte Pseudo-Likelihood-Version nach McCullagh und Nelder

Auch für die restringierte Pseudo-Likelihood-Version nach McCullagh und Nelder muss für die theoretischen Überlegungen mit einer Quasi-Verteilung gearbeitet werden. Im Gegensatz zu den EQL-Versionen kann für die PL^M jedoch das Integral über die Exponentialfunktion der PL^M gelöst werden:

$$\begin{aligned}
& \int_{-\infty}^{\infty} \exp(PL^M(\mu_i, \varphi_i, \lambda, \beta | y_i)) dy_i \\
&= \int_{-\infty}^{\infty} \exp \left\{ -\frac{n-L-1}{n} \frac{1}{2} \log(2\pi\varphi_i V(\mu_i)) - \frac{1}{2} \frac{r_P^2(y_i, \mu_i)}{\varphi_i} \right\} \\
&= \int_{-\infty}^{\infty} \left(\frac{1}{\sqrt{2\pi}\sqrt{\varphi_i V(\mu_i)}} \right)^{\frac{n-L-1}{n}} \exp \left\{ -\frac{1}{2} \frac{r_P^2(y_i, \mu_i)}{\varphi_i} \right\} dy_i \\
&= \left(\frac{1}{\sqrt{2\pi}\sqrt{\varphi_i V(\mu_i)}} \right)^{-\frac{L+1}{n}} \underbrace{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sqrt{\varphi_i V(\mu_i)}} \exp \left\{ -\frac{1}{2} \frac{(y_i - \mu_i)^2}{\varphi_i V(\mu_i)} \right\} dy_i}_{\text{Dichte einer } N(\mu_i, \varphi_i V(\mu_i))} \\
&= \left(\frac{1}{\sqrt{2\pi}\sqrt{\varphi_i V(\mu_i)}} \right)^{-\frac{L+1}{n}}
\end{aligned}$$

Der gesuchte Vorfaktor ist also $(2\pi\varphi_i V(\mu_i))^{-\frac{L+1}{2n}}$ und somit ist die Quasi-Dichte f_Q^M der $PL^M(\mu_i, \varphi_i, \lambda, \beta | y_i)$ gegeben durch

$$f_Q^M(y_i | \mu_i, \varphi_i, \lambda, \beta) = (2\pi\varphi_i V(\mu_i))^{-\frac{L+1}{2n}} \exp(PL^M(\mu_i, \varphi_i, \lambda, \beta | y_i)).$$

Die sich ergebende Quasi-Dichte für den kompletten Beobachtungsvektor \mathbf{y} lautet

$$f_Q^M(\mathbf{y} | \mu, \varphi, \lambda, \beta) = \prod_{i=1}^n \left\{ (2\pi\varphi_i V(\mu_i))^{-\frac{L+1}{2n}} \exp(PL^M(\mu_i, \varphi_i, \lambda, \beta | y_i)) \right\}$$

und entspricht der Dichte einer $N(\boldsymbol{\mu}, \text{diag}(\varphi_i V(\mu_i)))$.

Somit besitzt die PL^M keine echte Dichte aber obige Quasi-Dichte $f_Q^M(\mathbf{y})$. Die zu dieser Dichte gehörende Log-Likelihood wird im folgenden Quasi-Pseudo-Likelihood nach McCullagh und Nelder genannt und mit QPL^M bezeichnet. Sie lässt sich schreiben als

$$\begin{aligned} QPL^M(\boldsymbol{\mu}, \boldsymbol{\varphi}, \lambda, \beta | \mathbf{y}) &= -\frac{1}{2} \frac{L+1}{n} \sum_{i=1}^n \log(2\pi\varphi_i V(\mu_i)) + PL^M(\boldsymbol{\mu}, \boldsymbol{\varphi}, \lambda, \beta | \mathbf{y}) \\ &= -\frac{1}{2} \frac{L+1}{n} \sum_{i=1}^n \log(2\pi\varphi_i V(\mu_i)) - \frac{1}{2} \frac{n-L-1}{n} \sum_{i=1}^n \log(2\pi\varphi_i V(\mu_i)) - \sum_{i=1}^n \frac{1}{2} \frac{r_P^2(y_i, \mu_i)}{\varphi_i} \\ &= PL(\boldsymbol{\mu}, \boldsymbol{\varphi}, \lambda, \beta | \mathbf{y}) \end{aligned}$$

bzw. in Abhängigkeit des interessierenden Parametervektors $\boldsymbol{\theta} = (\boldsymbol{\alpha}^\top, \boldsymbol{\gamma}^\top, \lambda, \beta)^\top$ als

$$\begin{aligned} QPL^M(\boldsymbol{\alpha}, \boldsymbol{\gamma}, \lambda, \beta | \mathbf{y}) &= PL(\boldsymbol{\alpha}, \boldsymbol{\gamma}, \lambda, \beta | \mathbf{y}) \\ &= -\frac{1}{2} \sum_{i=1}^n \left(\log(2\pi) + \mathbf{z}_i^\top \boldsymbol{\gamma} + 2\beta \log \mu_i + \frac{(y_i - \mu_i)^2}{\exp(\mathbf{z}_i^\top \boldsymbol{\gamma}) \mu_i^{2\beta}} \right), \end{aligned}$$

mit $\mu_i = g^{-1}(\mathbf{x}_i^\top \boldsymbol{\alpha})$. Somit entspricht die zur restringierten Pseudo-Likelihood nach McCullagh und Nelder gehörende Quasi-Likelihood der unrestringierten Pseudo-Likelihood. Für die PL^M tritt der schon oben erwähnte Fall ein, dass der Vorfaktor $(2\pi\varphi_i V(\mu_i))^{-\frac{L+1}{2n}}$ von den zu schätzenden Parametern abhängt. Um zu zeigen, dass die mit der PL^M gewonnenen Schätzer auch konsistent und asymptotisch normalverteilt und effizient sind, werden folgende drei Punkte bewiesen:

- (P1) Die mit der Quasi-Pseudo-Likelihood QPL^M gewonnenen Schätzer sind konsistent und asymptotisch normalverteilt und effizient.
- (P2) Die aus der QPL^M gewonnenen Schätzer sind asymptotisch gleich den aus der PL^M gewonnenen Schätzern.
- (P3) Die Eigenschaften der QPL^M -Schätzer übertragen sich asymptotisch auf die PL^M -Schätzer.

Folgende Behauptung beweist (P1).

Behauptung 1

Für die Einstellungen \mathbf{x}_i und \mathbf{z}_i der Lokations- und Dispersionseffekte seien die unabhängig verteilten Zufallsvariablen $y_i, i = 1, \dots, n$, mit Erwartungswert μ_i und Varianz $\varphi_i V(\mu_i)$ gegeben. Weiter gelte $g(\mu_i) = \mathbf{x}_i^\top \alpha$ und $\log(\varphi_i) = \mathbf{z}_i^\top \gamma$ für alle $i, i = 1, \dots, n$ und $g(\cdot)$ und $V(\cdot)$ wie in (2.19) auf Seite 39. Dann sind die durch die Maximierung der zugehörigen restringierten Quasi-Pseudolikelihood QPL^M gewonnenen Schätzer für $\theta = (\alpha^\top, \gamma^\top, \beta, \lambda)^\top$ konsistent und asymptotisch normalverteilt und effizient.

Beweis:

siehe Erdbrügge und Auer (2010), da die QPL^M der unrestringierten PL entspricht. q.e.d.

Als nächstes ist (P2) zu zeigen, d.h. dass die Maximierung der QPL^M zu den gleichen Schätzern für θ führt wie die Maximierung der PL^M . Dazu wird zunächst mit dem folgenden Lemma Behauptung 2 gezeigt, mit der dann wiederum (P2) bewiesen wird.

Lemma 1

Sei $h_{1n} : A \subset \mathbb{R}^k \rightarrow \mathbb{R}$ stetig für alle n und besitze ein globales Maximum bei $\hat{\mathbf{x}}_n$, d.h. $h_{1n}(\hat{\mathbf{x}}_n) > h_{1n}(\mathbf{x})$ für alle $\mathbf{x} \in A \setminus \{\hat{\mathbf{x}}_n\}$. Weiter seien für jedes $n \in \mathbb{N}$ folgende Bedingungen an h_{1n} erfüllt:

B(1) Für jedes genügend kleine $\delta > 0$ existiert ein $\xi > 0$, so dass $\forall \mathbf{x} \in A$ mit

$$|h_{1n}(\hat{\mathbf{x}}_n) - h_{1n}(\mathbf{x})| < \delta \text{ gilt: } \|\hat{\mathbf{x}}_n - \mathbf{x}\| < \xi.$$

B(2) Aus $\delta \rightarrow 0$ folgt $\xi \rightarrow 0$.

Sei $h_{2n} : A \rightarrow \mathbb{R}$ definiert durch $h_{2n}(\mathbf{x}) := h_{1n}(\mathbf{x}) + i_n(\mathbf{x})$, wobei für $i_n(\mathbf{x}) : A \rightarrow \mathbb{R}$ gilt: $\forall \varepsilon > 0 \exists n_\varepsilon$, so dass $|i_n(\mathbf{x})| < \varepsilon$ für alle $n \geq n_\varepsilon$ und für alle $\mathbf{x} \in A$. Dann gilt:

Die Maximalstelle $\tilde{\mathbf{x}}_n$ von h_{2n} und die Maximalstelle $\hat{\mathbf{x}}_n$ von h_{1n} konvergieren für $n \rightarrow \infty$ gegeneinander, d.h. $\forall \xi > 0$ existiert ein $n_\xi \in \mathbb{N}$ mit $\|\hat{\mathbf{x}}_n - \tilde{\mathbf{x}}_n\| < \xi$ für alle $n \geq n_\xi$.

Beweis:

Wähle o.B.d.A. $\varepsilon > 0$ genügend klein. Der Maximalwert von $h_{2n}(\mathbf{x})$ ist mindestens $h_{1n}(\hat{\mathbf{x}}_n) - \varepsilon$, da $h_{2n}(\mathbf{x}) \geq h_{1n}(\mathbf{x}) - \varepsilon$ für alle $\mathbf{x} \in A$ und für alle $n \geq n_\varepsilon$.

Die Menge \tilde{A}_n aller möglichen Maximalstellen für h_{2n} ist für alle $n \geq n_\varepsilon$ gegeben durch:

$$\begin{aligned} \tilde{A}_n &= \{\mathbf{x} \in A : h_{2n}(\mathbf{x}) \geq h_{1n}(\hat{\mathbf{x}}_n) - \varepsilon\} \\ &= \{\mathbf{x} \in A : h_{1n}(\mathbf{x}) + \underbrace{i_n(\mathbf{x})}_{\leq \varepsilon} \geq h_{1n}(\hat{\mathbf{x}}_n) - \varepsilon\} \\ &\in \{\mathbf{x} \in A : h_{1n}(\mathbf{x}) + \varepsilon \geq h_{1n}(\hat{\mathbf{x}}_n) - \varepsilon\} \\ &= \{\mathbf{x} \in A : h_{1n}(\hat{\mathbf{x}}_n) - h_{1n}(\mathbf{x}) \leq 2\varepsilon\} \\ &= \{\mathbf{x} \in A : \|\hat{\mathbf{x}}_n - \mathbf{x}\| \leq \xi\} \quad \text{wegen B(1) aus Lemma 1 und } \varepsilon \text{ genügend klein} \end{aligned}$$

Da $\forall n \in \mathbb{N}$ gilt: $\tilde{x}_n \in \tilde{A}_n$, folgt, dass $\|\hat{x}_n - \tilde{x}_n\| < \xi$. Mit B(2) aus Lemma 1 folgt die Behauptung. q.e.d.

Behauptung 2

Gegeben seien die Funktionenfolgen $f_{1n} : \mathbb{R}^n \rightarrow \mathbb{R}$ und $f_{2n} : \mathbb{R}^n \rightarrow \mathbb{R}$ mit $f_{2n}(\mathbf{x}) = f_{1n}(\mathbf{x}) + g_n(\mathbf{x})$. Es seien folgende Voraussetzungen erfüllt:

- (1) $\frac{f_{1n}(\mathbf{x})}{n}$ besitzt für jedes feste n ein globales Maximum bei $\hat{\mathbf{x}}_n$.
- (2) Für $\frac{f_{1n}(\mathbf{x})}{n}$ seien die Bedingungen B(1) und B(2) aus Lemma 1 erfüllt.
- (3) $\frac{g_n(\mathbf{x})}{n}$ konvergiert für n gegen unendlich gleichmäßig gegen Null, d.h. für jedes $\epsilon > 0$ existiert ein n_ϵ , so dass für alle $n > n_\epsilon$ gilt: $|\frac{g_n(\mathbf{x})}{n}| < \epsilon$ für alle $\mathbf{x} \in \mathbb{R}^n$.

Dann konvergieren die Maximalstelle $\hat{x}(f_{2n})$ von f_{2n} und die Maximalstelle $\hat{x}(f_{1n})$ von f_{1n} gegeneinander.

Beweis:

Sei nun $h_{2n}(x) := \frac{f_{2n}(x)}{n}$, $h_{1n}(x) := \frac{f_{1n}(x)}{n}$ und $i_n(x) = \frac{g_n(x)}{n}$. Dann gilt mit Lemma 1, dass die Maximalstelle $\tilde{\mathbf{x}}_n$ von $\frac{f_{2n}}{n}$ und die Maximalstelle $\hat{\mathbf{x}}_n$ von $\frac{f_{1n}}{n}$ für $n \rightarrow \infty$ gegeneinander konvergieren, d.h. für alle $\xi > 0$ existiert ein $n_\xi \in \mathbb{N}$ mit $\|\hat{\mathbf{x}}_n - \tilde{\mathbf{x}}_n\| < \xi$ für alle $n \geq n_\xi$. Da für festes $n \in \mathbb{N}$ gilt, dass $\operatorname{argmax}\left(\frac{f_{2n}(\mathbf{x})}{n}\right) = \operatorname{argmax}(f_{2n}(\mathbf{x}))$ und analog $\operatorname{argmax}\left(\frac{f_{1n}(\mathbf{x})}{n}\right) = \operatorname{argmax}(f_{1n}(\mathbf{x}))$ gilt die Behauptung. q.e.d.

Nach diesen Vorüberlegungen kann nun zumindest im Fall von unimodalen Verteilungen gezeigt werden, dass die Quasi-Schätzer und die restringierten Pseudolikelihood Schätzer gegeneinander konvergieren.

Behauptung 3

Für die Einstellungen \mathbf{x}_i und \mathbf{z}_i der Lokations- und Dispersionseffekte seien die unabhängig verteilten Zufallsvariablen $y_i, i = 1, \dots, n$, mit Erwartungswert μ_i und Varianz $\varphi_i V(\mu_i)$ gegeben. Weiter gelte $g(\mu_i) = \mathbf{x}_i^\top \alpha$ und $\log(\varphi_i) = \mathbf{z}_i^\top \gamma$ für alle $i, i = 1, \dots, n$ und $g(\cdot)$ und $V(\cdot)$ wie in (2.19) auf Seite 39. Weiter sei der Stichprobenraum \mathcal{X} von X beschränkt und es liegen orthogonale Designs in den Lokations- und Dispersionseffekten vor. Zusätzlich sei die zugrundeliegende Verteilung unimodal. Bezeichne $\theta = (\alpha^\top, \gamma^\top, \lambda, \beta)^\top$ den interessierenden Parametervektor, dann konvergieren für $n \rightarrow \infty$ die durch die Maximierung der QPL^M gewonnenen Schätzer θ_n^{QM} gegen die durch die Maximierung der PL^M gewonnenen Schätzer θ_n^M .

Beweis:

Zunächst wird der zusätzliche Term $c_n(\theta) := -\frac{1}{2} \frac{L+1}{n} \sum_{i=1}^n \log(2\pi\varphi_i V(\mu_i))$ betrachtet. Für die in dieser Arbeit unterstellte Varianzfunktion $V(\mu_i) = \mu_i^{2\beta}$ und die unterstellten Linkfunktionen

$h(\varphi_i) = \log(\varphi_i)$ und $g(\cdot)$, wobei g wie in (2.19), Seite 39, definiert, ergibt sich

$$\begin{aligned}
c_n(\theta) &= -\frac{1}{2} \frac{L+1}{n} \sum_{i=1}^n \log(2\pi \varphi_i V(\mu_i)) \\
&= -\frac{1}{2} \frac{L+1}{n} \sum_{i=1}^n \log\left(2\pi \exp(\mathbf{z}_i^\top \gamma) \mu_i^{2\beta}\right) \\
&= -\frac{1}{2} \frac{L+1}{n} \sum_{i=1}^n \left[\log(2\pi) + \mathbf{z}_i^\top \gamma + 2\beta \log \mu_i \right] \\
&= (L+1) \left[\underbrace{-\frac{1}{2} \log(2\pi)}_{=:T_1} - \underbrace{\frac{1}{2} \frac{1}{n} \sum_{i=1}^n \mathbf{z}_i^\top \gamma}_{=:T_{2n}} - \underbrace{\beta \frac{1}{n} \sum_{i=1}^n \log \mu_i}_{=:T_{3n}} \right] \\
&=: (L+1) \tilde{c}_n(\theta).
\end{aligned}$$

Der Term T_1 entspricht einer Konstanten. Da ein orthogonales Design vorliegt, gilt, dass $T_{2n} = \frac{1}{2} \frac{1}{n} \sum_{i=1}^n \mathbf{z}_i^\top \gamma = \frac{\gamma_0}{2}$ unabhängig von n ist. Unter der Annahme, dass die PL^M -Schätzer und die QPL^M -Schätzer gegen ein festes $\theta \in \Theta$ konvergieren, kann o.B.d.A. angenommen werden, dass der Parameterraum Θ beschränkt ist und somit auch T_{2n} . Für den Term T_{3n} ergibt sich

$$2\beta \frac{1}{n} \sum_{i=1}^n \log \mu_i = \begin{cases} 2\beta \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i^\top \alpha = \alpha_0 & : \lambda = 0 \\ \frac{2\beta}{\lambda} \frac{1}{n} \sum_{i=1}^n \log(\lambda \mathbf{x}_i^\top \alpha + 1) & : \lambda \neq 0. \end{cases} \quad (3.1)$$

Ist der Stichprobenraum \mathcal{X} von X beschränkt und gilt obige Annahme, ist auch T_{3n} für beliebiges $\theta \in \Theta$ beschränkt, d.h. $|T_{3n}(\theta)| \leq T_3$ für alle n . Dies bedeutet, dass auch für $\tilde{c}_n(\theta)$ gilt: $|\tilde{c}_n(\theta)| \leq \tilde{c}$, wobei \tilde{c} fest.

Mit $c_n(\theta) = (L+1)\tilde{c}_n(\theta)$ lässt sich die PL^M schreiben als

$$\begin{aligned}
PL^M(\alpha, \gamma, \lambda, \beta | \mathbf{y}) &= (n-L-1)\tilde{c}_n(\theta) - n \frac{1}{2} \frac{1}{n} \sum_{i=1}^n \frac{r_P^2(y_i, \mu_i)}{\varphi_i} \\
&= (n-L-1)\tilde{c}_n(\theta) - \frac{n}{2} r(\mu, \varphi | \mathbf{y})
\end{aligned}$$

Dabei kann $r(\mu, \varphi | \mathbf{y}) := \frac{1}{n} \sum_{i=1}^n \frac{r_P^2(y_i, \mu_i)}{\varphi_i}$ als mittleres skaliertes Pearson-Residuum aufgefasst werden.

Somit gilt für die QPL^M :

$$\begin{aligned} QPL^M(\alpha, \gamma, \lambda, \beta | \mathbf{y}) &= PL^M(\alpha, \gamma, \lambda, \beta | \mathbf{y}) + (L+1)\tilde{c}_n(\theta) \\ &= \underbrace{(n-L-1)\tilde{c}_n(\theta) - \frac{n}{2}r(\mu, \varphi | \mathbf{y})}_{:=f_{1n}(\theta)} + \underbrace{(L+1)\tilde{c}_n(\theta)}_{:=g_n(\theta)} \end{aligned}$$

Weiter gilt

$$\left| \frac{g_n(\theta)}{n} \right| = \frac{L+1}{n} |\tilde{c}_n(\theta)| \leq \frac{L+1}{n} \tilde{c} \xrightarrow{n \rightarrow \infty} 0.$$

Auf Grund der Beschränktheit von $g_n(\theta)$ konvergiert $\frac{g_n(\theta)}{n}$ sogar gleichmäßig gegen Null. Somit ist Voraussetzung (3) von Behauptung 2 erfüllt. Voraussetzung (2) ist durch die Unimodalität der zugrundeliegenden Verteilung erfüllt. Es bleibt Voraussetzung (1) zu zeigen, dass $\frac{f_{1n}(\theta)}{n}$ für festes L gegen eine von n unabhängige Funktion konvergiert, die ein eindeutiges Maximum besitzt.

$$\begin{aligned} \frac{f_{1n}(\theta)}{n} &= \frac{n-L-1}{n} \tilde{c}_n(\theta) - \frac{1}{2} r(\mu, \varphi | \mathbf{y}) \\ &\xrightarrow{n \rightarrow \infty} \tilde{c}_n(\theta) - \frac{1}{2} \frac{1}{n} \sum_{i=1}^n \frac{r_P^2(y_i, \mu_i)}{\varphi_i} \\ &= -\frac{1}{2} \frac{1}{n} \sum_{i=1}^n \log(2\pi\varphi_i V(\mu_i)) - \frac{1}{2} \frac{1}{n} \sum_{i=1}^n \frac{r_P^2(y_i, \mu_i)}{\varphi_i} \\ &= \frac{1}{n} \sum_{i=1}^n PL(\alpha, \gamma, \lambda, \beta | y_i) \\ &= \frac{1}{n} PL(\alpha, \gamma, \lambda, \beta | \mathbf{y}). \end{aligned}$$

$\frac{f_{1n}(\theta)}{n}$ konvergiert somit gegen die durchschnittliche PL , die als echte Log-Likelihood ein eindeutiges Maximum besitzt.

Somit konvergieren die Maximalstellen mit Behauptung 2 gegeneinander.

q.e.d.

Nach diesen Vorüberlegungen ist es nun möglich (P3) zu beweisen, nämlich dass sich die Eigenschaften der QPL^M -Schätzer auf die PL^M -Schätzer übertragen.

Satz 3

Sei $\mathbf{y} = (y_1, \dots, y_n)^\top$ der Beobachtungsvektor, $PL^M(\alpha, \gamma, \lambda, \beta | \mathbf{y})$ die restringierte Pseudolikelihood nach McCullagh und Nelder und $\theta_0 = (\alpha^\top, \gamma^\top, \beta, \lambda)^\top = (\theta_{01}, \dots, \theta_{0s})$ mit $s = 4 + |L| + |D|$ der wahre Parametervektor. Dann gilt:

i) Die durch die PL^M gewonnenen Schätzer θ_{nj}^M sind für alle $j, j = 1, \dots, s$, konsistent für den wahren Parameter θ_{0j} , d.h.

$$P(|\theta_{nj}^M - \theta_{0j}| > \xi) \xrightarrow{n \rightarrow \infty} 0 \quad \text{für alle } \xi > 0.$$

ii) θ_n^M ist asymptotisch normalverteilt und effizient, d.h.

$$\theta_n^M \xrightarrow{d} N\left(\theta_0, \frac{1}{n} \mathcal{I}(\theta_0)^{-1}\right).$$

Beweis:

(P1) Mit Behauptung 3 gilt

$$|\theta_{nj}^M - \theta_{nj}^{QM}| < \varepsilon \quad \text{für alle } n \geq n_{\varepsilon, \theta}. \quad (*)$$

Des Weiteren gilt mit Behauptung 1

$$P\left(|\theta_{nj}^{QM} - \theta_{0j}| > \delta\right) \xrightarrow{n \rightarrow \infty} 0 \quad \text{für alle } \delta > 0 \quad (**)$$

Wähle nun $\xi = \varepsilon + \delta$ mit $\varepsilon, \delta > 0$ beliebig, dann gilt $\forall n > n_{\varepsilon, \theta}$:

$$\begin{aligned} & P\left(|\theta_{nj}^M - \theta_{0j}| > \xi\right) \\ &= P\left(|\theta_{nj}^M - \theta_{nj}^{QM} + \theta_{nj}^{QM} - \theta_{0j}| > \xi\right) \\ &\leq P\left(\underbrace{|\theta_{nj}^M - \theta_{nj}^{QM}|}_{< \varepsilon(*)} + |\theta_{nj}^{QM} - \theta_{0j}| > \xi\right) \\ &\leq P\left(|\theta_{nj}^{QM} - \theta_{0j}| > \underbrace{\xi - \varepsilon}_{\delta}\right) \xrightarrow{n \rightarrow \infty} 0 \quad \text{mit } (**)$$

(P2) Mit Behauptung 1 gilt für den Schätzvektor θ_n^{QM} der Quasi-Pseudolikelihood

$$\theta_n^{QM} \xrightarrow{d} N\left(\theta_0, \frac{1}{n} \mathcal{I}(\theta_0)^{-1}\right) \quad (*),$$

und mit Behauptung 3 gilt auch

$$\theta_n^M - \theta_n^{QM} \xrightarrow{P} \mathbf{0}_s$$

Somit gilt mit dem Satz von Slutsky (siehe Serfling (1980) S.19 oder Lehmann (1983) S.

337):

$$\theta_n^{QM} + \theta_n^M - \theta_n^{QM} \xrightarrow{d} N\left(\theta_0, \frac{1}{n} \mathcal{I}(\theta_0)^{-1}\right) + \mathbf{0}_s$$

und somit

$$\theta_n^M \xrightarrow{d} N\left(\theta_0, \frac{1}{n} \mathcal{I}(\theta_0)^{-1}\right) \quad \text{q.e.d.}$$

Somit konnten also auch für die auf der restringierten Pseudo-Likelihood nach McCullagh und Nelder beruhenden Schätzer die für die PL-Schätzer geltenden asymptotischen Eigenschaften gezeigt werden.

3.3 Restringierte Pseudo-Likelihood Version nach Lee und Nelder

Auch für die restringierte Pseudo-Likelihood nach Lee und Nelder muss zuerst die Quasi-Verteilung bestimmt werden. Das Integral über die Exponentialfunktion der PL^L ergibt sich dabei zu

$$\begin{aligned}
& \int_{-\infty}^{\infty} \exp \{ PL^L(\mu_i, \varphi_i | y_i) \} dy_i \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi} \sqrt{\varphi_i V(\mu_i)}} \exp \left\{ -\frac{1}{2} \frac{r_P^2(y_i, \mu_i)}{\varphi_i(1-h_{ii})} \right\} dy_i \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\varphi_i V(\mu_i)}} \exp \left\{ -\frac{1}{2} \frac{(y_i - \mu_i)^2}{\varphi_i V(\mu_i)(1-h_{ii})} \right\} dy_i \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\varphi_i V(\mu_i)(1-h_{ii})}} \exp \left\{ -\frac{1}{2} \frac{(y_i - \mu_i)^2}{\varphi_i V(\mu_i)(1-h_{ii})} \right\} \sqrt{1-h_{ii}} dy_i \\
&= \sqrt{1-h_{ii}} \underbrace{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\varphi_i V(\mu_i)(1-h_{ii})}} \exp \left\{ -\frac{1}{2} \frac{(y_i - \mu_i)^2}{\varphi_i V(\mu_i)(1-h_{ii})} \right\} dy_i}_{\text{Dichte einer } N(\mu_i, \varphi_i V(\mu_i)(1-h_{ii}))} \\
&= \sqrt{1-h_{ii}}
\end{aligned}$$

Dabei ist $h_{ii} = (W^{1/2}DX(X^T DWDX)^{-1}X^T DW^{1/2})_{ii}$ mit

$$W = \text{diag} \left(\frac{1}{\varphi_i V(\mu_i)} \right) = \text{diag} \left(\frac{1}{\varphi_i \mu_i^{2\beta}} \right) \text{ und } D = \text{diag} \left(\frac{\partial \mu_i}{\partial \eta_i} \right) = \text{diag} (\mu_i^{1-\lambda}).$$

Der gesuchte Vorfaktor ist also $\frac{1}{\sqrt{1-h_{ii}}}$ und somit ist die Quasi-Dichte f^Q der $PL^L(\mu_i, \varphi_i | y_i)$ gegeben durch

$$f^Q(y_i | \mu_i, \varphi_i) = \frac{1}{\sqrt{1-h_{ii}}} \exp \{ PL^L(\mu_i, \varphi_i | y_i) \}$$

Die Quasi-Dichte des kompletten Beobachtungsvektors \mathbf{y} ergibt sich zu

$$f^Q(y_1, \dots, y_n | \mu, \varphi) = \prod_{i=1}^n \frac{1}{\sqrt{1-h_{ii}}} \exp \{ PL^L(\mu_i, \varphi_i | y_i) \}.$$

Dies entspricht der Dichte einer $N(\mu, \text{diag}(\varphi_i V(\mu_i)(1-h_{ii})))$.

Somit besitzt auch die PL^L keine echte Dichte aber obige Quasi-Dichte $f^Q(\mathbf{y})$. Die zu der Dichte $f^Q(\mathbf{y})$ gehörende Log-Likelihood wird im Folgenden Quasi-Pseudo-Likelihood genannt und mit QPL^L gekennzeichnet. Sie lässt sich schreiben als

$$\begin{aligned} QPL^L(\mu, \varphi | \mathbf{y}) &= -\frac{1}{2} \sum_{i=1}^n \log(1-h_{ii}) + PL^L(\mu, \varphi | \mathbf{y}) \\ &= \underbrace{-\frac{1}{2} \sum_{i=1}^n \log(1-h_{ii})}_{:=c(\theta)} - \frac{1}{2} \sum_{i=1}^n \log(2\pi\varphi_i V(\mu_i)) - \frac{1}{2} \sum_{i=1}^n \frac{r_P^2(y_i, \mu_i)}{\varphi_i(1-h_{ii})}. \end{aligned}$$

Es tritt also erneut der Fall ein, dass der Vorfaktor $\frac{1}{\sqrt{1-h_{ii}}}$ von den zu schätzenden Parametern abhängt. Für diesen Fall werden die asymptotischen Eigenschaften nur für den Parametervektor $\theta = (\alpha^\top, \gamma^\top)^\top$ nachgewiesen, d.h. für bekanntes λ und β . Dazu werden die Punkte (P1) bis (P4) gezeigt.

(P1) Die mit der Quasi-Pseudo-Likelihood QPL^L gewonnen Schätzer sind konsistent und asymptotisch normalverteilt und effizient.

- (P2) Der Term $-\frac{1}{2} \sum_{i=1}^n \log(1 - h_{ii})$ konvergiert für alle $i, i = 1, \dots, n$, gegen eine Konstante.
- (P3) Die aus der QPL^L gewonnenen Schätzer sind asymptotisch gleich den aus der PL^L gewonnenen Schätzern.
- (P4) Die Eigenschaften der QPL^L -Schätzer übertragen sich asymptotisch auf die PL^L -Schätzer.

Zuerst werden die theoretischen Eigenschaften der QPL^L -Schätzer für den Spezialfall von bekannten λ und β mit Hilfe von Satz 2 nachgewiesen.

Behauptung 4

Für die Einstellungen \mathbf{x}_i und \mathbf{z}_i der Lokations- und Dispersionseffekte seien die unabhängig verteilten Zufallsvariablen $y_i, i = 1, \dots, n$, mit Erwartungswert μ_i und Varianz $\varphi_i V(\mu_i)$ gegeben. Weiter gelte $g(\mu_i) = \mathbf{x}_i^\top \alpha$ und $\log(\varphi_i) = \mathbf{z}_i^\top \gamma$ für alle $i, i = 1, \dots, n$, mit $g(\cdot)$ und $V(\cdot)$ wie in (2.19). Dabei seien λ und β bekannt. Dann sind die durch die Maximierung der zugehörigen restringierten Quasi-Pseudolikelihood QPL^L gewonnen Schätzer für $\theta = (\alpha^\top, \gamma^\top)^\top$ konsistent und asymptotisch normalverteilt und effizient.

Beweis:

mit Satz 2, Nachweis der Regularitätsbedingungen siehe Anhang A.2

q.e.d.

Als nächstes wird gezeigt, dass der Term $-\frac{1}{2} \sum_{i=1}^n \log(1 - h_{ii})$ für $n \rightarrow \infty$ gegen $\frac{1}{2} \text{rg}(X)$ konvergiert, wenn für das Maximum der h_{ii} gilt, dass es kleiner als $\frac{c}{n^k}$ ist, wobei $k \in (\frac{1}{2}, \infty)$.

Behauptung 5

Es sei folgende Voraussetzung erfüllt:

(V) Es gibt ein $n_1 \in \mathbb{N}$, so dass für alle $n \geq n_1$ gilt: $\max_{i=1, \dots, n}(h_{ii}) \leq \frac{c}{n^k}$, wobei $k \in (\frac{1}{2}, \infty)$ fest.

Dann konvergiert $-\frac{1}{2} \sum_{i=1}^n \log(1 - h_{ii})$ für $n \rightarrow \infty$ gegen $\frac{1}{2} \text{rg}(X)$, d.h.

$$-\frac{1}{2} \sum_{i=1}^n \log(1 - h_{ii}) \xrightarrow{n \rightarrow \infty} \frac{1}{2} \text{rg}(X). \quad (3.2)$$

Dabei ist h_{ii} wie in Definition 10 definiert und $\text{rg}(X)$ bezeichne den Rang der Matrix X .

Beweis:

Für $\log(1 - z)$ gilt folgende Reihendarstellung:

$$\log(1 - z) = - \sum_{k=1}^{\infty} \frac{z^k}{k}, \quad \text{für } |z| \leq 1.$$

Da für alle $i, i = 1, \dots, n$, gilt $0 \leq h_{ii} \leq 1$, siehe Toutenburg (2003) S. 346 gilt:

$$\begin{aligned} -\frac{1}{2} \sum_{i=1}^n \log(1 - h_{ii}) &= \frac{1}{2} \sum_{i=1}^n \sum_{k=1}^{\infty} \frac{h_{ii}^k}{k} \\ &= \frac{1}{2} \sum_{k=1}^{\infty} \sum_{i=1}^n \frac{h_{ii}^k}{k}, \quad \text{da alle Summanden } > 0 \\ &= \frac{1}{2} \sum_{i=1}^n h_{ii} + \frac{1}{2} \sum_{k=2}^{\infty} \frac{1}{k} \sum_{i=1}^n h_{ii}^k \\ &= \frac{1}{2} \text{rg}(X) + \frac{1}{2} \sum_{i=1}^n \sum_{k=2}^{\infty} \frac{h_{ii}^k}{k}, \end{aligned}$$

da für alle n gilt: $\sum_{i=1}^n h_{ii} = \text{rg}(W^{1/2}DX) = \text{rg}(X)$, wenn $W^{1/2}D$ Vollrang hat. Dies ist gegeben,

da die Diagonaleinträge von D und W alle größer als Null sind.

Somit bleibt zu zeigen, dass $\frac{1}{2} \sum_{k=2}^{\infty} \frac{1}{k} \sum_{i=1}^n h_{ii}^k$ für $n \rightarrow \infty$ gegen Null konvergiert. Bezeichne dazu $\max_{i=1, \dots, n}(h_{ii}) =: \max(h_{ii})$.

Für alle $n \geq n_1$ gilt:

$$\begin{aligned}
& \frac{1}{2} \sum_{i=1}^n \sum_{k=2}^{\infty} \frac{1}{k} h_{ii}^k \\
& \leq \frac{1}{4} \sum_{i=1}^n \sum_{k=2}^{\infty} h_{ii}^k \\
& \leq \frac{1}{4} \sum_{i=1}^n \sum_{k=2}^{\infty} \max(h_{ii})^k \\
& = \frac{1}{4} n \left(\sum_{k=0}^{\infty} \max(h_{ii})^k - 1 - \max(h_{ii}) \right) \\
& = \frac{n}{4} \left(\frac{1}{1 - \max(h_{ii})} - 1 - \max(h_{ii}) \right) \\
& = \frac{n}{4} \cdot \frac{1 - (1 - \max(h_{ii})) - \max(h_{ii})(1 - \max(h_{ii}))}{1 - \max(h_{ii})} \\
& = \frac{n}{4} \cdot \frac{[\max(h_{ii})]^2}{1 - \max(h_{ii})} \\
& \stackrel{(V)}{\leq} \frac{1}{4} \cdot \underbrace{\frac{1}{1 - \max(h_{ii})}}_{\xrightarrow{n \rightarrow \infty} 1} \cdot \underbrace{\frac{c^2}{n^{2k-1}}}_{\xrightarrow{n \rightarrow \infty} 0} \xrightarrow{n \rightarrow \infty} 0
\end{aligned}$$

q.e.d.

Die folgende Behauptung zeigt, dass Voraussetzung (V) von Behauptung 5 erfüllt ist, wenn ein wiederholtes Design vorliegt.

Behauptung 6

Gegeben seien die Versuchspläne $X \in \mathbb{R}^{n \times L+1}$ und $Z \in \mathbb{R}^{n \times D+1}$, die jeweils aus m Wiederholungen der Ausgangspläne $X_A \in \mathbb{R}^{n_A \times L+1}$ und $Z_A \in \mathbb{R}^{n_A \times D+1}$ bestehen. D.h. jede der

n_A Zeilen der Ausgangspläne wird m -mal wiederholt. Dann folgt, dass $h_{ii} \leq \frac{1}{m}$ für alle $i, i = 1, \dots, n$.

Beweis:

Jede gegebene Einstellung \mathbf{x}_i der Lokationsfaktoren kommt m -mal vor. Da D und W nur von den Einstellungen \mathbf{x}_i und \mathbf{z}_i der Einflussfaktoren auf die Lokation und Dispersion und dem zu schätzendem Parametervektor $\theta = (\alpha^\top, \gamma^\top, \beta, \lambda)^\top$ abhängen, also unabhängig von \mathbf{y} sind, wird auch jede Zeile von $W^{1/2}DX$ m -mal wiederholt. Es folgt mit Toutenburg (2003) S.349, dass $h_{ii} \leq \frac{1}{m} = \frac{a}{n}$ für alle $i, i = 1, \dots, n$. q.e.d.

Unter der Annahme, dass nur wenige Effekte aktiv sind, kann ein faktorieller und sogar ein fraktioniert faktorieller Plan zu einem wiederholten Plan in den aktiven Effekten werden. Somit ist obige Behauptung auch für ursprünglich nichtwiederholte Pläne relevant.

Nun wird der Punkt (P3) gezeigt, dass die QPL^M -Schätzer gegen die PL^M -Schätzer konvergieren .

Behauptung 7

Gegeben seien die unabhängigen Zufallsvariablen $y_i, i = 1, \dots, n$, mit Erwartungswert μ_i und Varianz $\varphi_i V(\mu_i)$. Dann konvergieren die durch die Maximierung der QPL^L gewonnenen Schätzer θ_n^{QL} für $n \rightarrow \infty$ gegen die durch die Maximierung der PL^L gewonnenen Schätzer θ_n^L , wenn die Erhöhung von n zu wiederholten Plänen führt.

Beweis: Mit Behauptung 5 gilt:

$$\begin{aligned} QPL^L &= -\frac{1}{2} \sum_{i=1}^n \log(1 - h_{ii}) + PL^L(\mu, \varphi | \mathbf{y}) \\ &\stackrel{n \rightarrow \infty}{=} \frac{1}{2} \text{rg}(X) + PL^L(\mu, \varphi | \mathbf{y}) \end{aligned}$$

Somit konvergieren die Maximalstellen von QPL^L gegen die von PL^L .

q.e.d.

Nun können somit zumindest für bekannte λ und β die gewünschten Eigenschaften nachgewiesen werden.

Satz 4

Sei $\mathbf{y} = (y_1, \dots, y_n)^\top$ der Beobachtungsvektor, $PL^L(\mu, \varphi | \mathbf{y})$ die restringierte Pseudolikehood nach Lee und Nelder und $\theta_0 = (\alpha^\top, \gamma^\top)^\top = (\theta_{01}, \dots, \theta_{0s})$ mit $s = 2 + |L| + |D|$ der wahre Parametervektor. λ und β seien also bekannt. Dann gilt:

i) Die durch die PL^L gewonnenen Schätzer θ_{nj}^M sind für alle $j, j = 1, \dots, s$, konsistent für den wahren Parameter θ_{0j} , d.h.

$$P(|\theta_{nj}^L - \theta_{0j}| > \xi) \xrightarrow{n \rightarrow \infty} 0 \quad \text{für alle } \xi > 0.$$

ii) θ_n^L ist asymptotisch normalverteilt und effizient, d.h.

$$\theta_n^L \xrightarrow{d} N\left(\theta_0, \frac{1}{n} \mathcal{I}(\theta_0)^{-1}\right).$$

Beweis: analog zu Satz 3

Kapitel 4

Iteratives Verfahren zur Identifizierung aktiver Lokations- und Dispersionseffekte

Bis jetzt wurde immer davon ausgegangen, dass die aktiven Lokations- und Dispersionseffekte bekannt sind. Dies wird in den meisten Anwendungen jedoch nicht der Fall sein.

Für die Identifikation aktiver Lokationseffekte stehen viele Methoden, wie z.B. der Halb-Normal-Plot, vgl. Toutenburg et al. (1998), die Methode von Lenth (1989) oder die von Dong (1993) zur Verfügung.

Für die Identifikation von Dispersionseffekten existieren bis jetzt hauptsächlich nichtiterative Verfahren. Eine gängige und sehr effiziente Methode, siehe Fuller und Bisgaard (1995),

ist die von Bergman und Hynén (1997), die auf der Normalverteilungsannahme beruht. McGrath und Lin (2001) bemerken, dass sie jedoch das Signifikanzniveau nicht mehr einhält, wenn mehr als ein aktiver Dispersionseffekt vorhanden ist. Erscheint die Normalverteilungsannahme als nicht gerechtfertigt, kann die nichtparameterische Variante von McGrath und Lin (2002) verwendet werden, die jedoch antikonservativ ist.

Auer und Kunert (2006) weisen in ihrem Artikel darauf hin, dass es wichtig ist, bei Verfahren zur Identifikation von aktiven Lokationseffekten die Vermischungsstruktur des gegebenen Designs zu beachten, da es sonst leicht zu falschen Vorgehensweisen führen kann wie in Raghavarao und Altan (2003). Somit ist es naheliegend auch die Vermischungsstruktur zwischen Lokations- und Dispersionseffekten bei der Identifikation aktiver Effekte zu beachten. McGrath und Lin (2001) untersuchen diese Vermischung und stellen fest, dass zwei nichtidentifizierte aktive Lokationseffekte entweder zu fälschlicherweise identifizierten Dispersionseffekten oder auch zu nicht erkannten aktiven Dispersionseffekten führen können.

Als Beispiel möge folgendes Szenario gelten: Den Daten unterliege das Lokationsmodell $E(y) = 1.0A + 0.5B - 0.5AB$. Des Weiteren habe A einen Einfluss auf die Dispersion. μ bezeichne den Erwartungswert von y unter dem korrekten Modell und μ_A den Erwartungswert von y unter Vernachlässigung der Lokationseffekte B und AB . Des Weiteren sei Var1 die Tendenz der Varianz von y , wenn A einen positiven, und Var2 die Tendenz der Varianz von y , wenn A einen negativen Dispersionseffekt besitzt. Somit ergibt sich das in Tabelle 4 dargestellte Szenario.

Es stellt sich nun die Frage, was passiert, wenn nur der Lokationseffekt A erkannt wird,

Tabelle 4.1: Beispielszenario

A	B	AB	μ	μ_A	Var1	Var2
+	+	+	1	1	↑	↓
+	-	-	1	1	↑	↓
-	+	-	0	-1	↓	↑
-	-	+	-2	-1	↓	↑

also sowohl der Effekt von B als auch die Interaktion von A und B nicht ins Modell mit aufgenommen werden.

Situation 1:

Besitzt A einen positiven Einfluss auf die Varianz (Var1), wird dieser womöglich nicht mehr erkannt, da das Modell nur eine insgesamt große Streuung in den Daten erkennt.

Situation 2:

Besitzt A hingegen einen negativen Einfluss auf die Varianz (Var2), so wird dieser durch das falsche Lokationsmodell verstärkt. Ist A nicht aktiv, könnte in dieser Situation der Dispersionseffekt von A trotzdem fälschlicherweise als aktiv erkannt werden.

Aufgrund dessen schlägt McGrath (2003) einen Test vor, der überprüft, ob identifizierte Dispersionseffekte wirklich aktiv sind oder der Effekt fälschlicherweise durch zwei nicht identifizierte Lokationseffekte hervorgerufen wird. Er versucht also Fehlentscheidungen auf Grund von Situation 2 vorzubeugen.

Eine alternative Methode zur Identifizierung von aktiven Lokations- und Dispersionseffekten wird von Pan und Taam (2002) vorgestellt. Auch sie bedenken in ihrem Verfahren die mögliche Vermischung zwischen einem Dispersionsfaktor und einem Paar von Lokationseffekten. Sie stellen jedoch in Simulationstudien fest, dass die Auswirkungen in Situation 1 des Beispielszenarios in der Regel gravierender sind, weil häufig der Dispersionseffekt unentdeckt bleibt. Situation 2 hingegen führt nur selten zu fälschlicherweise als aktiv erkannten Dispersionseffekten und nur zu nicht zu gravierenden Effektvergrößerungen, wenn die Schätzer auf Basis der EQL, also mit Hilfe von Iteratively Weighted Least Squares (IWLS, siehe Seite 30), geschätzt werden und die Identifikation mit Hilfe der zugehörigen t -Statistiken durchgeführt wird. Basierend auf diesen Erkenntnissen schlagen sie ein iteratives Verfahren zur Identifikation aktiver Lokations- und Dispersionseffekte basierend auf doppelt Generalisierten Linearen Modellen (dGLMs) vor, siehe S. 29 ff., welches in Situation 2 nicht identifizierte aktive Dispersionseffekte aufdecken kann.

Im Folgenden wird das Verfahren von Pan und Taam noch einmal ausführlich beschrieben. Danach wird ein alternatives Verfahren vorgeschlagen, welches auf den Devianzdifferenzen beruht. Dabei wird sich auf das erweiterte Box-Cox Modell (2.3) mit bekanntem Transformationsparameter λ beschränkt. Ist der Transformationsparameter bekannt, so kann im Weiteren der Einfachheit halber davon ausgegangen werden, dass die Beobachtungen y_i direkt einer Normalverteilung entstammen. Die resultierende Devianz \widetilde{Dev} entspricht dabei den Devianzen der EQL und der PL, wenn für das Lokationsmodell normalverteilte Daten, die natürliche Linkfunktion $g(\mu) = \mu$ und eine konstante Varianzfunktion $V(\mu) = 1$ angenommen wird.

4.1 Identifikation aktiver Lokations- und Dispersionseffekte nach Pan und Taam

Um auch die durch falsche Lokationsmodelle in Situation 1 unerkannt gebliebenen Dispersionseffekte identifizieren zu können, schlagen Pan und Taam (2002) ein iteratives Vorgehen zur Identifikation aktiver Lokations- und Dispersionseffekte vor. Zuerst werden mit einer zweifachen Rückwärtsselektion die aktiven Lokations- und Dispersionseffekte aus vorab bestimmten Mengen möglicher aktiver Effekte auf Basis von IWLS identifiziert. Danach wird überprüft, ob ein aktiver Dispersionseffekt aufgrund der Vermischungsstruktur von Situation 2 übersehen wurde. Das genaue Vorgehen erfolgt wie in den Schritten 1 bis 5 beschrieben:

Schritt 1: *Bestimmung der Menge $\mathcal{L}_1 \subset \mathcal{L}$ möglicher Lokationseffekte*

Mit Hilfe gängiger Screening Methoden, wie z.B dem Half-Normal Plot oder der Methode von Lenth wird unter Vernachlässigung möglicher Dispersionseffekte eine Menge $\mathcal{L}_1 \in \mathcal{L}$ möglicher Lokationseffekte bestimmt.

- Schritt 2: *Bestimmung der Menge $\mathcal{D}_1 \subset \mathcal{D}$ möglicher Dispersionseffekte*

Die Menge \mathcal{D}_1 möglicher Dispersionseffekte wird, wenn möglich, als Menge aller Haupteffekte festgelegt. Liegt ein hochfraktioniertes Design vor, kann z.B. die Methode von Bergman und Hynén oder die Methode von Box und Meyer verwendet werden, damit die Ausgangsmenge \mathcal{D}_1 nicht zu groß wird.

Schritt 3: *Reduzierung der Mengen \mathcal{L}_1 und \mathcal{D}_1*

Mit Hilfe von IWLS wird ein gewichtetes lineares Modell an die Beobachtungen und ein Gamma GLM mit log-Link an die quadrierten Residuen, also ein doppelt generalisiertes lineares Modell (dGLM) angepasst. Dabei werden alle Einflussgrößen aus den zuvor bestimmten Mengen \mathcal{L}_1 und \mathcal{D}_1 in die Modelle mit aufgenommen. Nun werden, basierend auf den zugehörigen t-Werten, alle nichtsignifikanten Effekte entfernt. Somit reduzieren sich die Mengen \mathcal{L}_1 und \mathcal{D}_1 zu \mathcal{L}_2 und \mathcal{D}_2 .

Schritt 4: *Reduzierung der Mengen \mathcal{L}_2 und \mathcal{D}_2*

Wiederholung von Schritt 3 ausgehend von den Mengen \mathcal{L}_2 und \mathcal{D}_2 und Erhalt der reduzierten Mengen \mathcal{L}_3 und \mathcal{D}_3 . Alle in \mathcal{L}_3 und \mathcal{D}_3 enthaltenen Effekte werden nun als aktiv angesehen.

Schritt 5: *Identifizierung übersehener Dispersionseffekte*

Folgendes Vorgehen wird für **jeden** Dispersionseffekt $D \in \mathcal{D} \setminus \mathcal{D}_3$ ausgeführt:

- (i) Es werden alle Paare von Spalten aus der Designmatrix X, deren komponentenweises Produkt die zum Dispersionseffekt D gehörende Spalte ergibt, bestimmt.
- (ii) Die Anzahl der $\frac{n}{2}$ Paare wird möglichst reduziert. Dabei werden alle Paare entfernt, die zu den kleinsten Lokationseffekten gehören. Des Weiteren werden alle Paare entfernt, die einen Effekt enthalten, der auf Grund von Hintergrundwissen oder da er zu einer Wechselwirkung höherer Ordnung gehört, als inaktiv angesehen werden kann. Bleiben, insbesondere bei hochfraktionierten Designs, immer noch zu viele Paare über, werden die, deren aufsummierte absolute Effekte zu den kleinsten Werten führen, entfernt. Die übrigen Paare werden in

der Menge \mathcal{L}_P zusammengefasst.

- (iii) Nun werden alle Effekte, die in den Mengen $\mathcal{L}_P \cup \mathcal{L}_3$ und $\{D\} \cup \mathcal{D}_3$ enthalten sind, in die Modelle aufgenommen und erneut ein dGLM angepasst. Ist der zu Dispersionseffekt D gehörende t-Wert jetzt signifikant, wird D als aktiv angenommen.

Die als aktiv identifizierten Faktoren für die Lokation sind somit durch die Menge \mathcal{L}_3 gegeben. Und die Menge der aktiven Dispersionsfaktoren ist durch die Menge \mathcal{D}_3 vereinigt mit den in Schritt 5 identifizierten Dispersionseffekten gegeben. Pan und Taam bemerken, dass dabei insbesondere die in Schritt 5 identifizierten Dispersionseffekte genau durch Bestätigungsexperimente zu überprüfen sind.

Anhand von Simulationsstudien zeigen Pan und Taam, dass ihre Methode zu einer deutlich größeren Sensitivität für die Dispersionseffekte in Situation 1 und zu einer leicht verbesserten Sensitivität in Situation 2 führt. Gleichzeitig werden die Fehlerraten für nicht aktive Dispersionseffekte nur in einem für Screening-Experimente akzeptablen Rahmen erhöht. Die Studien beschränken sich dabei auf den Fall von normalverteilten Fehlern mit natürlicher Linkfunktion $g(\mu) = \mu$ und konstanter Varianzfunktion $V(\mu) = 1$. Zur Bestimmung der Ausgangsmengen \mathcal{L}_1 und \mathcal{D}_1 verwenden sie die Methode von Lenth bzw. die von Bergman und Hynén.

Das Verfahren von Pan und Taam hat jedoch folgende Nachteile:

- Pan und Taam bemerken selbst, dass es bei einer Modellierung mit dGLMs zu unrealistischen Schätzern kommen kann, wenn die Anzahl der zu schätzenden Parameter

größer als $\frac{n}{2}$, also die halbe Anzahl der Beobachtungen, ist. Diese Grenze wird jedoch in den Schritten 1 und 2 und insbesondere in Schritt 5 bei hochfraktionierten Designs leicht erreicht.

- Schritt 5 muss für jeden Dispersionseffekt, der nicht in Schritt 4 als aktiv identifiziert wurde, durchgeführt werden. Da für diesen Schritt viele Überlegungen bezüglich der ins Modell aufzunehmenden Paare von Lokationseffekten nötig sind, ist dieser Schritt nicht so einfach zu formalisieren und auf Grund seiner Komplexität für Anwender nur bedingt geeignet.
- Zur Identifizierung werden neben der Methode der dGLM oft noch zwei weitere Verfahren benötigt, und zwar eines zur Identifizierung von aktiven Lokationseffekten und eines zur Identifizierung von aktiven Dispersionseffekten.
- Nach Schritt 5 wird nicht überprüft, ob die Kenntnis eines zusätzlichen Dispersionseffektes nicht zu neuen signifikanten Lokationseffekten führt.

4.2 Überlegungen zur Devianzdifferenz

Um ein Verfahren zur Identifikation aktiver Effekte basierend auf den Devianzen zu bestimmen, muss zunächst eine Statistik festgelegt werden, anhand derer die Identifikation aktiver Effekte erfolgen soll. Es ist bekannt, dass zwei Lokationsmodelle A und B, wobei B eine Erweiterung von A darstellt, basierend auf den Devianzen miteinander verglichen werden können, wenn die zugrunde liegende Verteilung zur Klasse der Exponentialfamilien gehört, also ein echtes GLM vorliegt. Wie schon in Abschnitt 2.2.2 bemerkt, ist die

Devianz selbst im Allgemeinen nicht χ^2 -verteilt und auch nicht unabhängig von α . Jedoch ist die Differenz der skalierten Devianzen asymptotisch χ_r^2 -verteilt, wobei die Anzahl r der Freiheitsgrade $|B| - |A|$ entspricht. Dabei ist $|A|$ ($|B|$) die Anzahl der in das Modell A (B) eingehenden Lokationseffekte. Wird jedoch das erweiterte Box-Cox Modell M_{BC} unterstellt, weicht die Verteilung der Devianzdifferenzen relativ stark von der χ^2 -Verteilung ab. Dies wird an Grafik 4.1 deutlich. Unterstellt wurde hierbei ein Modell mit folgenden Modellen für den Erwartungswert μ und die Varianz $Var(y)$:

$$\mu = 2A + 1B - C + 0.5AB \quad \text{und}$$

$$Var(y) = \exp(1D).$$

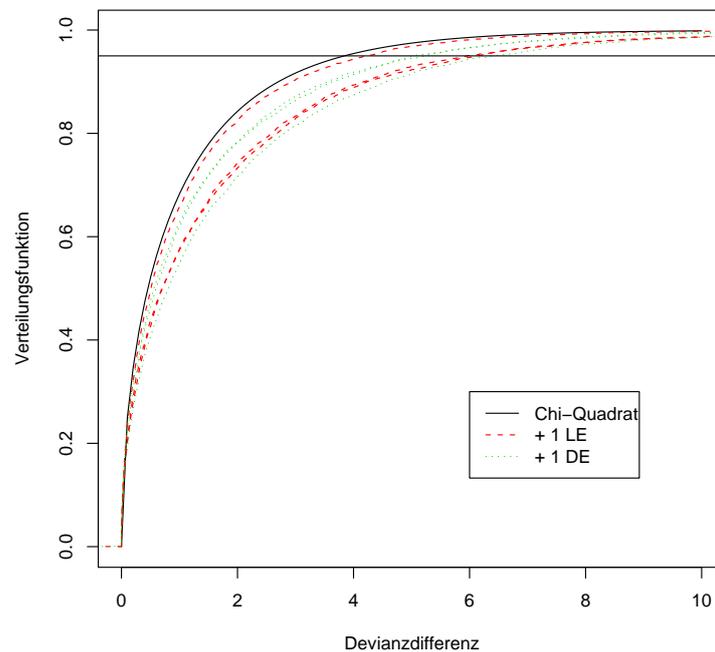


Abbildung 4.1: Verteilungsfunktion der Devianzdifferenzen für einen zusätzlichen Lokations- oder Dispersionseffekt (je 3x) im Vergleich zu einer χ_1^2 -Verteilung

Die Grafik enthält neben der Verteilungsfunktion einer χ_1^2 -Verteilung die empirischen Verteilungsfunktionen der Devianzdifferenzen zwischen dem wahren Modell und 3 Modellen mit einem zusätzlichen inaktiven Lokationseffekt (gestrichelte Linien) bzw. 3 Modellen mit einem zusätzlichen inaktiven Dispersionseffekt (gepunktete Linien). Es ist zu sehen, dass die aus 5000 Datensätzen berechneten empirischen Verteilungsfunktionen stark schwanken. Aber alle empirischen Verteilungsfunktionen führen unter diesen Modellvorstellungen zu deutlich größeren Quantilen als bei Unterstellung einer χ_1^2 -Verteilung. Somit wäre ein Test, der die Devianzdifferenzen mit den Quantilen einer χ_1^2 -Verteilung vergleichen würde antikonservativ.

Durch Simulationen werden globale 80%-, 90%- und 95%-Quantile für ein 2^4 -Design und ein 2^5 -Design aus 2000 bzw. 1000 Datensätzen bestimmt. Wie schon in der obigen Grafik gesehen, werden diese Quantile noch stark schwanken. Aufgrund der langen Rechenzeiten musste jedoch von einer Datensatzgröße von 10 000 oder größer abgesehen werden. Um ein globales Niveau bei der Identifizierung aktiver Effekte einzuhalten, muss die Verteilung der maximalen Devianzdifferenzen simuliert werden. D.h. es wird die Verteilung der maximalen Differenz zwischen der Devianz des korrekten Modells und der Devianz eines Modells mit einem zusätzlichen inaktiven Effekt simuliert. Dabei werden zusätzliche Lokations- und Dispersionseffekte getrennt betrachtet. In beiden Fällen werden jedoch alle Wechselwirkungen, auch die höherer Ordnung berücksichtigt, da bei fraktionierten Plänen gerade die Wechselwirkungen hoher Ordnung mit Haupteffekten vermischt werden. Des Weiteren werden die Quantile für verschiedene Anzahlen an aktiven Lokations- und Dispersionseffekten bestimmt, um eventuelle Abhängigkeiten zur Größe der unterstellten

Modelle erkennen zu können. Unter der Annahme, dass es meist nur sehr wenige aktive Dispersionseffekte und auch nur wenige aktive Lokationseffekte gibt, wurde sich dabei auf maximal 2 aktive Dispersionseffekte und maximal 6 aktive Lokationseffekte beschränkt. Somit können also bis zu 2 Dispersionseffekte und 7 Lokationseffekte identifiziert werden.

Tabelle 4.2: simulierte globale Quantile für die Lokationseffekte (1000 Datensätzen, 2^5 -Design)

Anzahl LE	Anzahl DE (Niveau)								
	0			1			2		
	0.8	0.9	0.95	0.8	0.9	0.95	0.8	0.9	0.95
0	8.0	9.3	10.8	8.2	9.6	11.3	8.5	10.1	11.9
1	8.1	9.5	11.0	8.3	9.7	11.5	8.5	10.4	11.9
2	8.3	9.9	11.7	8.6	9.8	11.5	9.0	10.9	12.5
3	8.5	10.2	11.6	9.5	11.1	13.2	9.4	11.6	13.0
4	8.7	10.1	11.5	9.9	11.7	13.4	10.4	12.5	14.5
5	9.1	10.8	12.4	10.3	12.4	14.6	12.0	15.2	17.2
6	9.1	11.0	12.8	10.6	12.5	15.1	11.7	14.4	17.0

Tabelle 4.2 enthält die simulierten globalen Quantile für die Identifizierung eines aktiven Lokationseffektes in einem 2^5 -Design. Es ist zu sehen, dass die Quantile von der Anzahl der aktiven Lokationseffekte abhängen. Es gilt, dass die Quantile um so größer werden, je größer die Anzahl der aktiven Lokationseffekte ist. Ebenso ist ein positiver Zusammenhang zwischen der Anzahl der aktiven Dispersionseffekte und der Quantilsgröße feststellbar. D.h. auch die lokalen Quantile wachsen mit der Anzahl der aktiven Faktoren, denn für die globalen Quantile wird bei steigender Anzahl an Einflussgrößen das Maximum über immer weniger Devianzdifferenzen gebildet. Würde die Verteilung der Devianzdifferenzen konstant sein, sollten die lokalen Quantile gleich bleiben und die globalen Quantile mit steigender Anzahl der aktiven Faktoren sinken. Diese Erkenntnis bestätigt sich, wenn die

entsprechenden lokalen Quantile betrachtet werden, siehe Anhang C, Tabelle C.1. Des Weiteren bestätigen diese Quantile auch die Erkenntnis aus Grafik 4.1, dass die Quantile nicht denen eine χ_1^2 -Verteilung entsprechen. Deren lokale 80%-, 90%- und 95%-Quantile lauten nämlich 1.64, 2.71 und 3.84 und sind deutlich kleiner als die entsprechenden simulierten Quantile.

Tabelle 4.3: simulierte globale Quantile für die Dispersionseffekte (1000 Datensätzen, 2^5 -Design)

Anzahl LE	Anzahl DE (Niveau)								
	0			1			2		
	0.8	0.9	0.95	0.8	0.9	0.95	0.8	0.9	0.95
0	7.7	9.4	11.1	7.6	9.0	10.4	7.7	9.3	10.7
1	8.4	10.1	11.8	8.4	10.2	11.7	8.5	10.5	12.0
2	9.3	11.3	12.7	9.0	10.6	12.3	8.7	10.9	12.8
3	10.1	11.8	13.7	9.7	11.3	13.1	9.9	11.8	13.7
4	11.2	13.1	15.3	11.1	13.3	15.3	10.9	13.3	15.5
5	12.6	14.6	16.7	11.9	15.0	16.9	12.5	15.0	17.5
6	14.0	16.8	18.9	13.4	16.1	18.4	14.1	16.3	18.7

Tabelle 4.3 enthält die globalen Quantile zur Identifizierung eines aktiven Dispersionseffektes in einem 2^5 -Design. Es ist festzustellen, dass es hier wiederum eine positive Abhängigkeit zwischen der Anzahl der aktiven Lokationseffekte und der Quantilsgröße gibt. Hingegen scheint es jetzt keinen Zusammenhang zwischen der Anzahl der aktiven Dispersionseffekte und der Quantilsgröße zu geben.

Die globalen Quantile für ein 2^4 -Design sind den Tabellen C.2 und C.3 in Anhang C zu entnehmen. Das Verhalten der Quantile entspricht denen für ein 2^5 -Design. Sie be-

ginnen jedoch bei kleineren Werten und enden mit viel größeren, haben also eine größere Spannweite.

Um zu überprüfen, ob die oben berechneten Quantile allgemeine Gültigkeit besitzen, ist zu überprüfen, ob sie unabhängig von der Effektgröße sind, d.h. ob die Quantile konstant bleiben, wenn in obiger Simulation statt einer jeweiligen Effektgröße von 1 eine von nur 0.5 oder eine von 2 gewählt worden wäre. Sie müssen also auf Skaleninvarianz geprüft werden.

Die folgenden Überlegungen können beweisen, dass die Devianzdifferenz unabhängig von α ist, aber nicht, dass sie auch unabhängig von γ ist. Jedoch machen sie plausibel, dass auch die Effektgröße von γ keine Rolle spielt.

Von Interesse ist das Verhalten der Devianzdifferenz, wenn α und γ unbekannt sind und somit ihre Schätzwerte einfließen. Dazu soll die Devianz zunächst in Matrixschreibweise dargestellt werden. Dabei wird die Kovarianzmatrix $\text{diag}(\exp \mathbf{z}_i^T \gamma)$ mit Σ_γ bezeichnet. Unter Ausnutzung der Tatsache, dass sich $\sum_{i=1}^n \mathbf{x}_i^T \gamma$ bei einem orthogonalen Versuchsplan zu $n\gamma_0$ ergibt, lässt sich die Devianz darstellen als

$$\widetilde{Dev}_{BC}(\alpha, \gamma | \lambda, \mathbf{y}) = \log(2\pi) + n\gamma_0 + (\mathbf{y} - X\alpha)^T \Sigma_{-\gamma} (\mathbf{y} - X\alpha).$$

Wird nun für α der gewichtete KQ-Schätzer $(X^T \Sigma_{-\gamma} X)^{-1} X^T \Sigma_{-\gamma/2} \mathbf{y}$ eingesetzt, ist zu

sehen, dass die Devianz nur noch von γ abhängt:

$$\begin{aligned}
\widetilde{Dev}_{BC}(\alpha, \gamma | \lambda, \mathbf{y}) &\geq \widetilde{Dev}_{BC}(\gamma | \hat{\alpha}, \lambda, \mathbf{y}) \\
&= \log(2\pi) + n\gamma_0 + (\mathbf{y} - \mathbf{X}\hat{\alpha})^\top \Sigma_{-\gamma} (\mathbf{y} - \mathbf{X}\hat{\alpha}) \\
&= \log(2\pi) + n\gamma_0 + (\Sigma_{-\gamma/2} \mathbf{y})^\top \omega^\perp (\Sigma_{-\gamma/2} \mathbf{X}) (\Sigma_{-\gamma/2} \mathbf{y}) \\
&= \log(2\pi) + n\gamma_0 + (\Sigma_{-\gamma/2} \mathbf{e})^\top \omega^\perp (\Sigma_{-\gamma/2} \mathbf{X}) (\Sigma_{-\gamma/2} \mathbf{e})
\end{aligned}$$

Sei nun γ^* der wahre Wert, so dass $\mathbf{e} = D_{\gamma^*/2} \varepsilon$, wobei $\varepsilon \sim N(0, \sigma^2 I_n)$. Dann ist

$$\widetilde{Dev}_{BC}(\gamma | \hat{\alpha}, \lambda, \mathbf{y}) = \log(2\pi) + n\gamma_0 + \varepsilon^\top \Sigma_{\gamma^*/2} \Sigma_{-\gamma/2} \omega^\perp (\Sigma_{-\gamma/2} \mathbf{X}) \Sigma_{-\gamma/2} \Sigma_{\gamma^*/2} \varepsilon$$

Somit ergibt sich $\hat{\gamma}$ zu

$$\begin{aligned}
\hat{\gamma} &= \operatorname{argmin}_{\gamma \in \Gamma} \left\{ \varepsilon^\top \Sigma_{\gamma^*/2} \Sigma_{-\gamma/2} \omega^\perp (\Sigma_{-\gamma/2} \mathbf{X}) \Sigma_{-\gamma/2} \Sigma_{\gamma^*/2} \varepsilon : \gamma \in \Gamma \right\} \\
&= \operatorname{argmin}_{\gamma \in \Gamma} \left\{ \varepsilon^\top \Sigma_{(\gamma^* - \gamma)/2} \omega^\perp (\Sigma_{(\gamma^* - \gamma)/2} \tilde{\mathbf{X}}) \Sigma_{(\gamma^* - \gamma)/2} \varepsilon : \gamma \in \Gamma \right\} \\
&= \operatorname{argmin}_{\gamma \in \Gamma} \underbrace{\left\{ \varepsilon^\top \Sigma_{\gamma/2} \omega^\perp (\Sigma_{\gamma/2} \tilde{\mathbf{X}}) \Sigma_{\gamma/2} \varepsilon : \gamma \in \Gamma \right\}}_{\exists \varepsilon^\top \omega^\perp (\tilde{\mathbf{X}}) \varepsilon \sim \sigma^2 \chi_{n-r}^2}
\end{aligned}$$

Somit enthält die zu minimierende Menge zumindest ein Element, welches unabhängig vom wahren Parameter γ^* ist.

Gehen wir davon aus, dass die wahre Kovarianzmatrix $\Sigma_{\gamma_1 + \gamma_2}$ ist und die Dispersionsmodelle B und A mit $\Sigma_B = \Sigma_{\hat{\gamma}_1 + \hat{\gamma}_2}$ und $\Sigma_A = \Sigma_{\tilde{\gamma}_1}$ miteinander verglichen werden sollen, dann

ergibt sich die Devianzdifferenz zu

$$\begin{aligned}
DD_{B-A} &:= \widetilde{Dev}_{BC}(\hat{\gamma} | \hat{\alpha}, \lambda, \mathbf{y}) - \widetilde{Dev}_{BC}(\tilde{\gamma} | \tilde{\alpha}, \lambda, \mathbf{y}) \\
&= \varepsilon^T \Sigma_{(\gamma_1+\gamma_2)/2} \Sigma_{-(\hat{\gamma}_1+\hat{\gamma}_2)/2} \omega^\perp \left(\Sigma_{-(\hat{\gamma}_1+\hat{\gamma}_2)/2} \mathbf{X} \right) \Sigma_{-(\hat{\gamma}_1+\hat{\gamma}_2)/2} \Sigma_{(\gamma_1+\gamma_2)/2} \varepsilon \\
&\quad - \varepsilon^T \Sigma_{(\gamma_1+\gamma_2)/2} \Sigma_{-\tilde{\gamma}_1/2} \omega^\perp \left(\Sigma_{-\tilde{\gamma}_1/2} \mathbf{X} \right) \Sigma_{-\tilde{\gamma}_1/2} \Sigma_{(\gamma_1+\gamma_2)/2} \varepsilon
\end{aligned}$$

Unter der Annahme, dass $\gamma_2 = 0$ ist und die Schätzer $\hat{\gamma}_1, \hat{\gamma}_2, \tilde{\gamma}_1$ gegen die wahren Werte von γ_1 und γ_2 konvergieren, so konvergiert die Devianzdifferenz gegen 0 und ist somit unabhängig von der Effektgröße γ_1 . Die Devianz der Box-Cox Erweiterung stimmt, wie schon oben erwähnt, für bekanntes λ mit der auf der PL basierten Devianz überein. Und für die mit Hilfe der PL bestimmten Schätzer konnte in Erdbrügge und Auer (2010) die Konsistenz der Schätzer nachgewiesen werden.

Sollen die Lokationsmodelle $LM_A : \xi = X_1 \alpha_1$ und $LM_B : \xi = [X_1, X_2] \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix}$ für gegebenes Dispersionsmodell mit Kovarianzmatrix Σ_γ miteinander verglichen werden, so ergibt sich die Devianzdifferenz zu

$$\begin{aligned}
DD_{B-A} &:= \widetilde{Dev}_{BC}(\hat{\gamma} | \hat{\alpha}, \lambda, \mathbf{y}) - \widetilde{Dev}_{BC}(\tilde{\gamma} | \tilde{\alpha}, \lambda, \mathbf{y}) \\
&= \varepsilon^T \Sigma_{\gamma/2} \Sigma_{-\hat{\gamma}/2} \omega^\perp \left(\Sigma_{-\hat{\gamma}/2} [X_1, X_2] \right) \Sigma_{-\hat{\gamma}/2} \Sigma_{\gamma/2} \varepsilon \\
&\quad - \varepsilon^T \Sigma_{\gamma/2} \Sigma_{-\tilde{\gamma}/2} \omega^\perp \left(\Sigma_{-\tilde{\gamma}/2} X_1 \right) \Sigma_{-\tilde{\gamma}/2} \Sigma_{\gamma/2} \varepsilon
\end{aligned}$$

Somit sollte auch in diesem Falle die Devianzdifferenz zumindest asymptotisch unabhängig von γ sein, wenn sowohl $\tilde{\gamma}$ als auch $\hat{\gamma}$ für $\alpha_2 = 0$ gegen den wahren Parameter konvergieren.

Da obige Überlegungen es nur plausibel erscheinen lassen, dass die Devianzdifferenz unabhängig von γ ist, werden nun einige der Simulationen zur Bestimmung der globalen Quantile, die mit einer Effektgröße von 1 durchgeführt wurden, mit den Effektgrößen 0.5 und 2 wiederholt. Außerdem beruht die Simulation dieses Mal auf 2000 Datensätzen um sicherere Aussagen treffen zu können. Die Tabellen 4.4 und 4.5 geben die globalen Quantile für die verschiedenen Werte von γ wieder. Wie den Tabellen zu entnehmen ist, unterscheiden sich diese Quantile nur relativ gering von den zuvor bestimmten. Eine eindeutige Abhängigkeit der Quantile von der Effektgröße ist nicht erkennbar. Nur bei den Quantilen für die Lokationseffekte ist eine leichte Tendenz zu einem negativen Zusammenhang sichtbar.

Tabelle 4.4: simulierte globale Quantile für die Lokationseffekte für unterschiedliche Effektgrößen von γ (2^5 -Design, 2000 Daten)

LE	DE	$\gamma=0.5$			$\gamma=1$			$\gamma=2$		
		0.8	0.9	0.95	0.8	0.9	0.95	0.8	0.9	0.95
3	1	9.4	11.2	13.1	9.2	10.8	12.6	8.8	10.5	12.2
3	2	10.1	12.2	14.0	9.7	11.9	14.4	8.8	10.8	13.1
4	1	9.8	11.7	13.6	9.8	11.7	13.8	9.2	11.3	13.1
4	2	10.8	12.8	14.6	10.5	12.7	14.7	10.1	12.3	14.7

Werden Konfidenzintervalle für die 95%-Quantile bestimmt, überschneiden sie sich für die Effektgrößen 0.5 und 2 von γ stark, siehe Tabelle 4.6. Die Konfidenzintervalle für einen zusätzlichen Lokationseffekt bei 3 aktiven Lokationseffekten und einem aktiven Dispersionsseffekt überdecken sich sogar. Somit geben diese Simulationen keinen Hinweis darauf, dass die 95%-Quantile der Devianzdifferenz von γ abhängen. (Zumindest nicht für die hier betrachteten Effektgrößen, die aber eine sinnvolle Spannweite der Varianz abdecken.)

Tabelle 4.5: simulierte globale Quantile für die Dispersionseffekte für unterschiedliche Effektgrößen von γ (2^5 -Design, 2000 Daten)

LE	DE	$\gamma=0.5$			$\gamma=1$			$\gamma=2$		
		0.8	0.9	0.95	0.8	0.9	0.95	0.8	0.9	0.95
3	1	9.9	11.8	13.7	9.9	11.9	14.1	9.8	11.8	13.6
3	2	10.3	12.4	14.1	10.0	12.0	13.8	10.0	12.0	13.8
4	1	11.0	13.1	15.1	11.1	13.2	15.1	10.9	12.9	15.0
4	2	11.2	13.6	15.5	11.3	13.4	15.6	11.0	13.1	15.4

Tabelle 4.6: Konfidenzintervalle für 95%-Quantile (2^5 -Design, 10*200 Daten)

LE	DE	Lokation		Dispersion	
		$\gamma=0.5$	$\gamma=2$	$\gamma=0.5$	$\gamma=2$
3	1	[12.36, 13.12]	[11.87, 14.39]	[13.73, 14.42]	[13.13, 14.31]
3	2	[12.93, 14.38]	[12.27, 13.11]	[12.93, 14.38]	[12.27, 13.11]
4	1	[12.71, 13.58]	[12.45, 13.43]	[12.72, 13.58]	[12.45, 13.43]
4	2	[13.86, 15.07]	[13.47, 14.38]	[13.86, 15.07]	[13.47, 14.38]

Somit können nun zwei Modelle A und B , wobei B einen zusätzlichen Effekt enthält, auf Basis der zugehörigen Devianzdifferenz $\widetilde{Dev}_B - \widetilde{Dev}_A$ miteinander verglichen werden. Ist die Devianzdifferenz größer als das entsprechende simulierte Quantil, kann davon ausgegangen werden, dass der in Modell B enthaltene zusätzliche Effekt aktiv ist.

4.3 Identifikation aktiver Lokations- und Dispersionseffekte beruhend auf den Devianzdifferenzen

Aufgrund der Überlegungen im vorangegangenen Abschnitt wird nun folgendes auf den Devianzdifferenzen beruhendes iteratives Verfahren zur Identifikation aktiver Lokations- und Dispersionseffekte vorgeschlagen:

Schritt 1: *Bestimmung aktiver Lokationseffekte*

Zuerst werden unter Vernachlässigung möglicher Dispersionseffekte schrittweise mit Hilfe der Devianzdifferenzen aktive Lokationseffekte identifiziert. Die Vernachlässigung eventueller Dispersionseffekte sollte die Identifizierung aktiver Lokationseffekte in Situation 2 erleichtern.

Schritt 2: *Bestimmung eines aktiven Dispersionseffektes*

Gegeben das Lokationsmodell aus Schritt 1, wird anhand der Devianzdifferenz überprüft, ob es einen aktiven Dispersionseffekt gibt. Der größte aktive Effekt wird in das Modell aufgenommen.

Schritt 3: *Iterative Verbesserung der Modelle*

Wurde in Schritt 2 ein Dispersionseffekt identifiziert, wird, gegeben die jeweiligen aktuellen Modelle, abwechselnd überprüft, ob ein weiterer Lokations- bzw. Dispersionseffekt aufgenommen werden muss, bis keine neuen Effekte mehr aufgenommen werden. Die Menge der aktiven Lokations- bzw. Dispersionseffekte nach diesem Schritt wird mit \mathcal{L}_0 bzw. \mathcal{D}_0 bezeichnet.

Schritt 4: Identifikation übersehener Dispersionseffekte

Um Situation 1 aufzufangen, wird nun für **jeden** nicht identifizierten Dispersionseffekt $D \in \mathcal{D} \setminus \mathcal{D}_0$ wie folgt vorgegangen:

Gegeben die Lokations- und Dispersionseffekte aus Schritt 3 und zusätzlich Dispersionseffekt D wird mit der Devianzdifferenz überprüft, ob mindestens ein zusätzlicher Lokationseffekt signifikant ist. Ist dies der Fall und gilt zusätzlich, dass der Dispersionsfaktor D unter diesem neuen Modell signifikant ist, werden er und die zusätzlich identifizierten Lokationseffekte mit ins Modell aufgenommen.

Die Vorteile dieses Verfahrens im Vergleich zur Methode von Pan und Taam sind leicht ersichtlich:

- Die komplette Identifikation aktiver Effekte beruht auf einer einzigen Methode.
- Die Gefahr von zu großen Modellen ist durch die Vorwärtsselektion gering.
- Die Identifikation der aktiven Dispersionseffekte mit Hilfe der Devianzdifferenzen kann im Gegensatz zur Methode von Bergman und Hynén auch bei nicht normalverteilten Daten angewendet werden.
- Der Schritt zur Identifikation übersehener Dispersionseffekte ist viel einfacher als der entsprechende Schritt in der Methode von Pan und Taam.
- Schritt 4 führt nicht nur zu einer Verbesserung des Dispersions- sondern auch zu einer des Lokationsmodells.

Diese Vorteile sind natürlich nur dann relevant, wenn die auf den Devianzdifferenzen basierende Methode eine mindestens ähnlich hohe Sensitivität und eine ähnlich kleine Fehlerrate aufweist, wie die Methode von Pan und Taam. Dies soll im nächsten Kapitel anhand von Simulationsstudien untersucht werden.

Der Nachteil des auf den Devianzdifferenzen beruhenden Verfahrens liegt darin, dass die exakte Verteilung der Devianzdifferenzen nicht bekannt ist und die Quantile somit in Abhängigkeit der Designgröße und der Anzahl der aktiven Lokations- und Dispersionseffekte simuliert werden müssen.

4.4 Vergleich der iterativen Verfahren

In diesem Abschnitt soll mit Hilfe von Simulationsstudien überprüft werden, ob das vorgeschlagene iterative Verfahren zu ähnlich guten Werten führt, wie das von Pan und Taam. Vorab wird auch untersucht, ob die Identifikation aktiver Lokationseffekte unter Verwendung der Devianzdifferenzen eine mindestens so große Sensitivität aufweist, wie die Identifikation mit Hilfe von Standardmethoden, wenn mögliche Dispersionseffekte vernachlässigt werden. Gegebenenfalls wäre sonst Schritt 1 zu modifizieren. In allen folgenden Simulationsstudien sollen sowohl einfache Situationen mit wenigen großen aktiven Effekten als auch schwierige Situationen mit relativ vielen aktiven Effekten betrachtet werden.

Für die schwierigen Situationen wird in Analogie zu Pan und Taam (2002) ein 2^4 -Design für die Faktoren A bis D betrachtet. Die den Simulationsstudien zugrunde liegenden

Lokations- und Dispersionsmodelle lauten dabei wie folgt:

Lokationsmodell $LM_0 : \mu = 0$

Lokationsmodell $LM_1 : \mu = 0 + 0.5A + 1.0B + 2.0C + 0.5AB$

Lokationsmodell $LM_2 : \mu = 0 + 0.5A + 1.0B + 2.0C - 0.5AB$

Dispersionsmodell $DM_0 : \sigma^2 = \exp(0)$

Dispersionsmodell $DM_1 : \sigma^2 = \exp(0 + 0.5A + 1.0B)$

Dispersionsmodell $DM_2 : \sigma^2 = \exp(0 + 0.5A - 1.0B)$

Dispersionsmodell $DM_3 : \sigma^2 = \exp(0 + 1.0D)$.

Für die einfache Situation wird ein 2^5 -Design für die Faktoren A bis E betrachtet. Dabei werden folgende Modelle untersucht:

Lokationsmodell $LM_0 : \mu = 0$

Lokationsmodell $LM_1 : \mu = 0 + 2A + 1.0B + 1AB$

Lokationsmodell $LM_2 : \mu = 0 + 2A + 1.0B - 1AB$

Dispersionsmodell $DM_0 : \sigma^2 = \exp(0)$

Dispersionsmodell $DM_1 : \sigma^2 = \exp(0 + 1.0B)$

Dispersionsmodell $DM_2 : \sigma^2 = \exp(0 - 1.0B)$

Dispersionsmodell $DM_3 : \sigma^2 = \exp(0 + 1.0D)$.

Werden die Lokationsmodelle LM_1 und LM_2 mit den Dispersionsmodellen DM_1 und DM_2 kombiniert ist der Dispersionseffekt B jeweils mit den Lokationseffekten A und AB vermischt. Die Kombinationen LM_1, DM_1 und LM_2, DM_2 entsprechen dabei Situation 2 aus dem Beispielszenario am Anfang dieses Kapitels und die Kombinationen LM_1, DM_2 und LM_2, DM_1 entsprechen Situation 1. Mit Dispersionsmodell DM_3 soll zusätzlich eine

Situation betrachtet werden, in der der Dispersionsfaktor nicht im Lokationsmodell, auch nicht über Wechselwirkungen mit anderen Faktoren, enthalten ist.

Zuerst wird mit Hilfe von Simulationen überprüft, ob mit Hilfe der Devianzdifferenzen die aktiven Lokationseffekte unter Vernachlässigung eventueller Dispersionseffekte mindestens genauso gut identifiziert werden können wie mit den Methoden von Lenth (1989) und Dong (1993). Die Methode von Lenth wird betrachtet, da Pan und Taam sie verwenden. Da die Methode von Dong jedoch eine höhere Sensitivität bei wenigen aktiven Effekten aufweist, siehe Kunert (1997), soll auch sie betrachtet werden. Beide bestehenden Verfahren beruhen auf dem Median basierten Varianzschätzer s_0 , der wie folgt definiert ist:

$$s_0 = \frac{3}{2} \operatorname{med}_{p=1, \dots, L} (|\hat{\alpha}_p|).$$

Dabei sind die $\hat{\alpha}_p$ die ungewichteten KQ-Schätzer des vollen Modells. Lenth definiert nun den Pseudo Standard Error (PSE) als

$$PSE = \frac{3}{2} \operatorname{med}_{\substack{|\hat{\alpha}_p| \leq 2.5s_0 \\ p=1, \dots, L}} (|\hat{\alpha}_p|) \quad (4.1)$$

und erklärt die Effekte $p \in \mathcal{L}$ zum lokalen ζ -Niveau als aktiv, für die gilt

$$\left| \frac{\hat{\alpha}_p}{PSE} \right| > t \left(\frac{1 + \zeta}{2}, \frac{L}{3} \right).$$

Dabei ist $t(\zeta, r)$ das ζ -Quantil einer t-Verteilung mit r Freiheitsgraden. Soll ein globales ζ -Niveau eingehalten werden, so ist das $\frac{1+\zeta}{2}$ - Quantil der t-Verteilung mit $\frac{L}{3}$ Freiheitsgraden zu verwenden.

Ein alternativer Schätzer für die Standardabweichung ist der Adaptive Standard Error

(ASE) von Dong (1993), der wie folgt definiert ist:

$$ASE = \sqrt{\frac{1.08}{m_L} \sum_{\substack{|\hat{\alpha}_p| \leq 2.56s_0, \\ p=1, \dots, L}} \hat{\alpha}_p^2}, \tag{4.2}$$

wobei m_L die Anzahl der Effektschätzer ist, die betragsmäßig kleiner als $2.56s_0$ sind. Die mit dem ASE standardisierten Koeffizientenschätzer werden dann mit den Quantilen einer t-Verteilung mit $0.69L$ Freiheitsgraden verglichen. Für die Begründung der Freiheitsgrade, siehe Kunert (1997).

Tabelle 4.7: Sensitivitätsvergleich zwischen ASE, PSE und DD (2^4 -Design, 1000 Datensätze)

	Modell	A	B	C	D	AB	AC	AD	BC	BD	CD	$\overset{A}{BC}$	$\overset{A}{BD}$	$\overset{A}{CD}$	$\overset{B}{CD}$	$\overset{AB}{CD}$	GFR
PSE	LM_0	0.2	0.1	0.2	0.2	0.1	0.1	0.4	0.2	0.2	0.2	0.1	0.3	0.2	0.2	0.4	1.4
ASE	DM_0	0.4	0.5	0.5	0.8	0.3	0.2	0.2	0.3	0.3	0.3	0.7	0.8	0.8	0.2	0.7	4.6
DD		0.4	0.2	0.2	0.4	0.4	0.3	0.0	0.4	0.2	0.4	0.3	0.3	0.3	0.0	0.4	4.1
PSE	LM_0	0.3	0.3	0.2	0.3	0.1	0.4	0.4	0.2	0.2	0.1	0.3	0.3	0.2	0.0	0.1	2.1
ASE	DM_1	0.8	0.8	0.3	0.8	0.5	0.9	1.0	0.5	0.5	0.4	1.0	0.5	0.3	0.5	0.5	5.4
DD		0.2	0.6	0.2	0.1	0.1	0.0	0.0	0.1	0.1	0.0	0.4	0.0	0.1	0.3	0.0	2.0
PSE	LM_0	0.6	0.8	0.6	0.6	1.0	0.4	0.5	0.7	0.9	0.3	0.5	0.9	0.2	0.2	0.3	2.9
ASE	DM_2	0.7	0.9	1.2	0.8	1.0	1.1	0.7	1.1	0.9	0.4	0.9	0.9	0.4	0.5	0.5	4.4
DD		0.3	0.0	0.1	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.5
PSE	LM_1	2.8	16.4	72.8	0.1	1.8	0.1	0.0	0.0	0.1	0.2	0.0	0.0	0.1	0.0	0.1	0.5
ASE	DM_0	4.9	37.6	97.5	0.2	4.5	0.1	0.0	0.0	0.1	0.2	0.1	0.1	0.1	0.1	0.0	1.0
DD		1.8	39.7	94.8	0.0	2.7	0.1	0.1	0.0	0.1	0.0	0.0	0.1	0.0	0.0	0.0	0.4
PSE	LM_1	1.2	7.9	43.4	0.1	0.8	0.1	0.1	0.1	0.2	0.0	0.1	0.1	0.1	0.0	0.2	0.7
ASE	DM_1	3.0	19.1	81.4	0.0	3.2	0.1	0.1	0.2	0.2	0.0	0.1	0.1	0.2	0.0	0.3	1.0
DD		0.2	18.8	79.7	0.0	0.4	0.0	0.1	0.2	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.4
PSE	LM_1	1.2	3.9	14.7	0.1	1.3	0.3	0.1	0.6	0.1	0.1	0.4	0.1	0.0	0.0	0.0	0.9
ASE	DM_2	3.0	6.9	39.0	0.0	3.2	0.4	0.0	0.8	0.2	0.0	0.8	0.2	0.0	0.0	0.0	1.2
DD		2.6	2.9	38.8	0.0	0.4	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2
PSE	LM_2	2.4	14.4	72.8	0.0	2.2	0.1	0.0	0.2	0.0	0.1	0.1	0.2	0.2	0.1	0.1	0.9
ASE	DM_0	4.9	35.5	96.6	0.3	3.7	0.1	0.0	0.1	0.0	0.1	0.1	0.1	0.2	0.1	0.3	1.2
DD		2.0	37.0	95.2	0.0	1.6	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1
PSE	LM_2	1.3	8.7	47.0	0.0	1.7	0.1	0.0	0.3	0.0	0.1	0.1	0.0	0.0	0.1	0.0	0.5
ASE	DM_1	4.2	21.5	80.9	0.0	4.5	0.3	0.0	0.2	0.0	0.3	0.1	0.0	0.1	0.4	0.2	0.9
DD		2.0	19.5	82.4	0.0	2.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
PSE	LM_2	1.2	3.9	14.4	0.0	1.5	0.2	0.1	0.2	0.0	0.1	0.4	0.0	0.0	0.2	0.2	0.9
ASE	DM_2	1.9	10.7	37.4	0.0	2.3	0.2	0.1	0.1	0.0	0.1	0.2	0.0	0.1	0.4	0.5	1.0
DD		0.6	8.6	36.5	0.0	0.4	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1

Tabelle 4.7 gibt für jeden Effekt die Wahrscheinlichkeit als aktiv identifiziert zu werden in Prozent an. Die Werte beruhen auf 1000 simulierten Daten eines 2^4 -Designs mit den angegebenen Lokations- und Dispersionsmodellen. Die Werte der in Wahrheit aktiven Effekte sind fett dargestellt und entsprechen der Sensitivität des jeweiligen Verfahrens.

Die zu den inaktiven Effekten gehörenden Werte geben die individuellen Fehlerraten an. Die letzte Spalte enthält die Gesamtfehlerrate (GFR). Dies ist die Wahrscheinlichkeit, dass mindestens ein Effekt fälschlicherweise identifiziert wurde. An ihr ist also abzulesen, ob das globale Signifikanzniveau eingehalten wurde. Es ist zu sehen, dass die auf den Devianzdifferenzen DD beruhende Methode genau wie die Methode von Dong eine deutliche Verbesserung im Vergleich zur Methode von Lenth darstellt. Die Methode von Dong ist jedoch leicht besser als die auf den Devianzdifferenzen beruhende, insbesondere für die relativ kleinen Effekte. Die GFR liegt bis auf eine Ausnahme für die Methode von Dong immer unter 5%. Das Signifikanzniveau wird dabei bei allen Verfahren längst nicht ausgenutzt und ist für die Devianzdifferenzen meist am geringsten.

In Tabelle 4.8 sind die entsprechenden Simulationsergebnisse für den Fall eines 2^5 -Designs angegeben. Dabei wird sich hier auf die Angabe der Sensitivität der in den Lokationsmodellen LM_1 und LM_2 aktiven Effekte beschränkt. Hier ist zu sehen, dass die Devianzdifferenzen bei geringerer GFR ähnlich gute Ergebnisse wie die ASE liefert. Nur für den Fall LM_0, DM_0 ist die GFR für die Devianzdifferenz größer als die für die ASE, sie ist sogar größer als 5% und hält somit für diese 1000 Datensätze das globale Niveau nicht ein.

Insgesamt zeigen die Simulationsergebnisse, dass es keinen zwingenden Grund gibt, die Identifikation aktiver Lokationseffekte in Schritt 1 nicht mit Hilfe der Devianzdifferenzen durchzuführen.

Somit wird nun das vorgeschlagene iterative Verfahren anhand von Simulationen getestet. Tabelle 4.9 gibt die Sensitivität bzw. die Fehlerrate für die einzelnen Effekte in Prozent wieder. Dabei wurden 1000 Datensätze eines 2^4 -Designs simuliert und unter Verwendung

Tabelle 4.8: Sensitivitätsvergleich zwischen ASE, PSE und DD (2^5 -Design, 1000 Datensätze)

	Modell	A	B	AB	GFR
PSE	LM_0, DM_0	0.2	0.2	0.2	3.2
ASE		0.2	0.3	0.4	5.0
DD		0.3	0.3	0.3	6.4
PSE	LM_0, DM_1	0.3	0.3	0.2	3.2
ASE		0.4	0.3	0.3	4.4
DD		0.3	0.2	0.1	3.6
PSE	LM_0, DM_2	0.6	0.3	0.3	4.5
ASE		0.3	0.1	0.2	5.1
DD		0.0	0.2	0.1	4.1
PSE	LM_1, DM_0	100.0	83.3	82.3	2.8
ASE		100.0	95.8	96.9	4.0
DD		100.0	96.2	97.8	3.3
PSE	LM_1, DM_1	100.0	59.0	58.6	2.8
ASE		100.0	79.2	78.8	3.3
DD		100.0	78.1	76.7	1.1
PSE	LM_1, DM_2	99.8	57.2	56.3	3.0
ASE		100.0	77.0	77.7	3.0
DD		100.0	79.9	78.9	0.9
PSE	LM_2, DM_0	100.0	82.9	82.9	2.4
ASE		100.0	95.9	97.1	3.0
DD		100.0	96.3	97.1	1.9
PSE	LM_2, DM_1	99.8	53.0	54.5	2.3
ASE		100.0	78.4	76.5	2.9
DD		100.0	80.5	78.0	0.9
PSE	LM_2, DM_2	99.6	57.0	56.4	3.5
ASE		100.0	78.9	79.6	3.6
DD		100.0	76.7	77.5	2.0
PSE	LM_1, DM_3	99.6	58.3	58.9	3.1
ASE		100.0	79.6	79.0	2.9
DD		100.0	81.7	81.1	1.4
PSE	LM_2, DM_3	99.0	55.7	56.9	2.5
ASE		100.0	76.6	78.4	2.9
DD		100.0	77.5	79.1	0.9

der globalen 95%-Quantile ausgewertet. Neben den Ergebnissen mit Schritt 4 des vorgeschlagenen Verfahrens sind zusätzlich auch die ohne Schritt 4 wiedergegeben. Für die Identifikation der Lokationseffekte werden alle Wechselwirkungen, auch die höherer Ordnung, berücksichtigt. Für die Identifikation aktiver Dispersionseffekte wird sich jedoch

Es ist zu sehen, dass die Sensitivität sowohl der Lokationseffekte als auch der Dispersionseffekte durch Schritt 4 deutlich zunimmt. So steigt z.B. in Modell LM_2, DM_1 die Sensitivität für den Lokationseffekt A von 3.5% auf 21.6% und die des Dispersionseffektes B von 0.2% auf 36.6%. Gleichzeitig wird die Fehlerrate nur leicht erhöht. Für die Lokationseffekte liegt sie jetzt bei maximal 2% und für die Dispersionseffekte bei maximal 5.6%, welches gerade für ein Screening akzeptable Werte sind. Die Sensitivität der mit dem Dispersionsfaktor B vermischten Lokationseffekte A und AB kann in den der Situation 2 entsprechenden Modellen LM_1, DM_1 und LM_2, DM_2 nur geringfügig erhöht werden. So steigt sie z.B. für Modell LM_2, DM_2 für den Lokationseffekt AB von 1.2% auf 3.9%. Vergleicht man die Werte ohne Schritt 4, also nach Schritt 3, mit denen nach Schritt 1 aus Tabelle 4.7, so ist für die Modelle LM_1, DM_1 und LM_2, DM_2 aus Situation 2 eine klare Verbesserung der Sensitivität für die Lokationseffekte zu erkennen, die nicht mit dem Dispersionseffekt B vermischt sind. Die Effizienzsteigerung beträgt mehr als 100%. Für die Modelle aus Situation 1 fällt die Verbesserung durch Schritt 3 wesentlich geringer aus. Die Sensitivität der mit dem Dispersionseffekt B vermischten Lokationseffekte kann in Situation 2 nicht wesentlich durch die Schritte 2 und 3 verbessert werden, da diese Lokationseffekte, wenn überhaupt, nur unter Vernachlässigung des vermischten Dispersionseffektes entdeckt werden können. Jedoch wird Dispersionseffekt D bei Vorliegen von Dispersionsmodell D_3 auch nach Schritt 4 nicht erkannt. In diesem Fall ist die Fehlerrate für Dispersionseffekte A und B sogar größer als die Sensitivität von D .

In Tabelle 4.10 finden sich die Simulationsergebnisse für die Sensitivität bei einem 2^5 -Design. Die Ergebnisse sind wieder in Prozent angegeben und beruhen auf 1000 simulier-

Tabelle 4.10: Sensitivitätsanalyse für das auf den Devianzdifferenzen beruhende Verfahren *mit* und *ohne* Schritt 4, (2^5 -Design, Verwendung von globalen 95%-Quantilen, 1000 Datensätze)

Modell	Lokation					Dispersion			
	A	B	AB	$max(FR)$	\overline{FR}	A	D	$max(FR)$	\overline{FR}
LM_0, DM_0 , mit	0.8	0.4	0.2	1.0	0.47	1.9	2.0	1.9	2.1
LM_0, DM_0 , ohne	0.4	0.3	0.0	0.4	0.18	0.1	0.2	0.2	0.8
LM_0, DM_1 , mit	1.0	0.2	0.6	1.0	0.49	73.4	1.3	2.4	1.65
LM_0, DM_1 , ohne	0.4	0.2	0.2	0.6	0.22	71.9	0.1	0.2	0.15
LM_0, DM_2 , mit	0.4	0.3	0.2	0.8	0.44	69.9	2.1	2.1	1.63
LM_0, DM_2 , ohne	0.2	0.1	0.2	0.5	0.23	69.2	0.5	0.5	0.2
LM_1, DM_0 , mit	100.0	97.7	98.0	0.9	0.48	1.8	3.4	3.4	2.2
LM_1, DM_0 , ohne	100.0	94.4	95.9	0.4	0.17	0.2	0.5	0.5	0.2
LM_1, DM_1 , mit	99.9	68.3	68.8	1.1	0.53	65.8	2.6	3.2	2.7
LM_1, DM_1 , ohne	99.9	67.8	67.8	0.3	0.16	64.6	0.2	0.6	0.3
LM_1, DM_2 , mit	100.0	84.8	85.1	0.8	0.44	56.4	3.8	3.8	3.2
LM_1, DM_2 , ohne	100.0	76.3	74.7	0.4	0.17	47.2	0.5	0.5	0.43
LM_2, DM_0 , mit	100.0	97.7	98.2	0.9	4.5	1.6	3.5	3.5	2.3
LM_2, DM_0 , ohne	100.0	95.1	96.5	0.4	0.17	0.1	0.4	0.4	0.14
LM_2, DM_1 , mit	100.0	84.4	84.0	1.2	0.48	62.5	3.1	4.3	3.4
LM_2, DM_1 , ohne	100.0	75.0	76.4	0.3	0.14	52.7	0.2	0.7	0.43
LM_2, DM_2 , mit	99.9	72.9	72.8	0.8	0.46	59.8	2.3	2.4	2.2
LM_2, DM_2 , ohne	99.9	72.5	72.5	0.5	0.19	59.2	0.7	0.7	0.43
LM_1, DM_3 , mit	100.0	93.3	93.3	1.2	0.59	2.7	58.1	3.8	2.85
LM_1, DM_3 , ohne	100.0	74.3	74.7	0.4	0.15	0.9	39.0	0.9	0.58
LM_2, DM_3 , mit	100.0	95.2	95.2	1.2	0.51	2.7	59.8	3.6	2.48
LM_2, DM_3 , ohne	100.0	75.7	76.8	0.4	0.16	1.3	28.6	1.3	0.63

ten Datensätzen. Bei der Angabe der Fehlerraten (FR) für die nichtaktiven Effekte wird sich auf die maximale Fehlerrate $max(FR)$ und die mittlere Fehlerrate \overline{FR} getrennt nach Lokationseffekten und Dispersionseffekten beschränkt. Die Werte für die Sensitivität sind erneut hervorgehoben. Für die Modelle aus Situation 2 (LM_1, DM_1 bzw. LM_2, DM_2) ist nun keine wesentliche Verbesserung durch Schritt 4 mehr erkennbar. Für die Modelle aus Situation 1 steigt die Sensitivität des Dispersionseffektes und der zugehörigen vermischten Lokationseffekte trotz der schon hohen Sensitivität aber noch um 10%. Die Modelle mit Dispersionsmodell DM_3 , also einem Dispersionseffekt der nicht im Lokationsmodell enthalten ist, können durch Schritt 4 jedoch im Gegensatz zu den Simulationsergebnissen für das 2^4 -Design deutlich verbessert werden. Sowohl die Sensitivität der Lokationseffekte B und AB als auch die des Dispersionseffektes D selbst wächst um ca. 20%. Ein Vergleich

mit Tabelle 4.8 zeigt, dass Schritt 2 und 3 des iterativen Verfahrens für keines der betrachteten Modelle eine große Verbesserung bringen. Erneut bleibt die Fehlerrate in einer akzeptablen Größenordnung, sie liegt immer unter 5%.

Somit führt das auf den Devianzdifferenzen basierende iterative Verfahren genau wie das Vorgehen von Pan und Taam zu einer verbesserten Identifikation aktiver Effekte bei gleichzeitiger geringer Erhöhung der Fehlerraten.

Um das auf den Devianzdifferenzen beruhende Verfahren mit dem von Pan und Taam vorgeschlagenem besser vergleichen zu können, werden einige der Simulationen für das 2^4 -Design mit globalen 80%-Niveaus wiederholt. Das globale Niveau wurde dabei so gewählt, dass die Fehlerraten für die nichtaktiven Effekte nun ungefähr den von Pan und Taam erhaltenen Fehlerraten entsprechen. Die Ergebnisse finden sich zusammen mit den Simulationsergebnissen von Pan und Taam in Tabelle 4.11. Da Pan und Taam Wechselwirkungen höherer Ordnung nicht betrachtet haben, fehlen die Fehlerraten für die 3-fach und 4-fach Wechselwirkungen für ihr Verfahren.

Wie schon erwähnt, entsprechen sich nun die Fehlerraten ungefähr. Auch die Sensitivität der Dispersionseffekte ist annähernd gleich. Jedoch ist die Sensitivität der Lokationseffekte bei dem Verfahren von Pan und Taam in einigen Fällen deutlich besser. Ein konkreter Vergleich ist jedoch nicht möglich, da Pan und Taam nicht angeben, welche Quantile sie in den jeweiligen Schritten verwenden. Sie scheinen jedoch ein relativ geringes lokales Signifikanzniveau verwendet zu haben. Dies führt jedoch schnell zu einer zu großen Anzahl aktiver Effekte in ihrem ersten Modell. Besonders wenn in fraktionierten Designs

Tabelle 4.11: Sensitivitätsvergleich zwischen Verfahren mit Devianzdifferenzen und Verfahren von Pan und Taam (2^4 -Design, globale 80%-Quantile, 1000 Datensätze)

Modell	Param.	A	B	C	D	AB	AC	AD	BC	BD	CD	$\frac{A}{BC}$	$\frac{A}{BD}$	$\frac{A}{CD}$	$\frac{B}{CD}$	$\frac{AB}{CD}$
LM_1, DM_1	α	10.1	50.2	99.8	5.1	10.7	4.9	6.3	5.4	5.8	4.9	4.4	5.6	6.2	6.0	7.7
DD	γ	25.9	53.4	8.1	15.3											
LM_1, DM_1	α	24.0	75.4	99.8	4.5	25.5	4.3	4.8	5.5	3.8	2.5					
Pan u. Taam	γ	25.4	53.4	10.1	12.5											
LM_1, DM_2	α	39.1	58.1	99.5	3.3	31.7	3.3	5.6	4.9	3.2	2.9	5.0	5.7	6.2	4.4	4.6
DD	γ	24.7	40.4	7.4	15.0											
LM_1, DM_2	α	35.3	77.6	100.0	4.2	35.2	4.7	3.2	5.3	4.2	3.9					
Pan u. Taam	γ	24.4	45.5	9.1	13.5											
LM_2, DM_1	α	35.8	58.4	98.0	4.2	35.8	4.3	6.1	5.2	5.4	4.3	6.7	5.1	6.3	5.6	5.6
DD	γ	20.7	44.7	6.3	14.1											
LM_2, DM_1	α	36.1	79.1	99.8	3.9	33.2	3.9	3.4	4.9	4.4	3.5					
Pan u. Taam	γ	24.2	46.1	8.3	12.2											
LM_2, DM_2	α	11.7	62.4	99.9	4.0	13.0	4.7	7.4	5.8	5.2	4.5	5.9	6.0	7.2	7.3	6.6
DD	γ	27.3	53.7	6.2	14.6											
LM_2, DM_2	α	29.9	78.6	99.9	3.5	30.0	3.4	3.4	5.1	4.1	4.5					
Pan u. Taam	γ	21.9	51.7	9.0	13.9											

auch noch Wechselwirkungen höherer Ordnung betrachtet werden müssen. Pan und Taam scheinen sich jedoch auf die Haupteffekte und 2-fach Interaktionen beschränkt zu haben.

4.5 Beispiel

Im Folgenden soll das auf den Devianzdifferenzen basierende Verfahren zur Identifikation aktiver Lokations- und Dispersionseffekte auf zwei Datensätzen, die im Rahmen des Projektes C6 „Methoden der statistischen Versuchsplanung zu Analyse und Optimierung von komplexen Umformprozessen am Beispiel des CNC-gesteuerten Drückens“ im Sonderforschungsbereich 475 erhoben wurden, angewandt werden. Die Daten entstammen einem Produktplan und einem Combined Array und wurden in einer über beide Pläne randomisierten Reihenfolge erhoben. Der Zweck dieser doppelten Datenerhebung war der Vergleich beider Designs bezüglich ihrer Vor- und Nachteile zur Identifikation aktiver Lokations-

und Dispersionsfaktoren. Nähere Einzelheiten und die Ergebnisse des Vergleichs sind in Kunert et al. (2007) nachzulesen.

Es wurden jeweils 6 Designfaktoren, im Folgenden A, B, C, D, E und F , und 3 Störgrößen, im Folgenden m, n und o , betrachtet. Beide Pläne hatten einen Umfang von 32 Läufen. Die betrachtete Zielgröße A_{20} misst die während des Umformprozesses entstehende Blech-Aufdickung bzw. Ausdünnung. Sie ist definiert als die normierte Fläche zwischen der Soll-Blechdicke und der tatsächlich realisierten Blechdicke im umgeformten Bereich des Werkstücks.

Wird das vorgeschlagene Verfahren unter Verwendung der globalen 95%-Quantile auf den Produktplan angewandt, so werden die Faktoren B, E, F und n als aktive Lokationseffekte identifiziert und der Faktor D als aktiver Dispersionseffekt. Dies entspricht genau den Ergebnissen der mit einem Halb-Normal-Plot durchgeführten Analyse aus Kunert et al. (2007).

Auch die für das Combined-Array identifizierten Faktoren bringen keine neuen Erkenntnisse. Erneut werden die Effekte AF, B und F als aktive Lokationseffekte identifiziert. Und genau wie in der Standardanalyse mit einem Halb-Normal-Plot gibt es Hinweise darauf, dass auch der Haupteffekt von n einen Einfluss auf die Lokation haben könnte und C einen auf die Dispersion.

Somit führt das neue Verfahren bei normalverteilten Daten zu den gleichen Ergebnissen wie die üblichen Standardverfahren. Auch das neue Verfahren kann somit die Widersprüche zwischen den beiden Designs nicht aufheben.

Kapitel 5

Zusammenfassung und Ausblick

Häufig ist es in technischen Prozessen z.B. aufgrund hoher Kosten und/oder zeitaufwändiger Prozesse nicht möglich, wiederholte Designs durchzuführen. Somit ist es oft nötig, anhand weniger Versuchspunkte sowohl Lokations- als auch Dispersionseffekte zu identifizieren. Darüber hinaus ist in technischen Prozessen oft noch ein Zusammenhang zwischen Erwartungswert und Varianz zu beobachten. Kapitel 2 beschreibt die gut bekannten Transformationsmodelle und die immer verbreiteteren Generalisierten Linearen Modelle (GLM). Dabei werden zunächst für beide Ansätze die Standardverfahren erläutert, die es jedoch alle nicht erlauben, den Transformationsparameter gleichzeitig mit den Lokations- und Dispersionseffekten zu schätzen. Dies ermöglichen jedoch die von Erdbrügge und Auer (2010) eingeführten Erweiterungen. Sowohl die auf GLMs als auch die auf Transformationsmodellen beruhende Erweiterung werden ausführlich dargestellt. Des Weiteren werden die für die Extended Quasi-Likelihood (EQL) vorgeschlagenen restringierten Varianten auf die Pseudo-Likelihood übertragen.

Erdbrügge und Auer zeigen, dass die Schätzer für die Lokations- und Dispersionseffekte und für den Transformationsparameter konsistent und asymptotisch effizient und normalverteilt sind, wenn die GLM-Methode in Verbindung mit der Pseudo-Likelihood (PL) verwendet wird. In Kapitel 3 wird versucht, diese wünschenswerten Eigenschaften nachzuweisen, wenn statt der PL die EQL oder restringierte Varianten verwendet werden. Da die EQL und ihre restringierten Varianten jedoch keine echte Dichte besitzen und es auch nicht möglich ist, durch Integration eine Quasi-Dichte zu bestimmen, kann über diese Schätzer keine Aussage gemacht werden. Für die restringierten Varianten der PL kann jedoch ein Vorfaktor bestimmt werden, so dass die restringierte PL der Log-Likelihood einer echten Dichte entspricht. Für beide restringierten Varianten hängt der Vorfaktor jedoch von den zu schätzenden Parametern ab. Trotzdem konnten die Eigenschaften auf die mit der restringierten PL nach McCullagh und Nelder bestimmten Schätzer übertragen werden und für die auf der restringierten PL nach Lee und Nelder basierten Schätzer zumindest für den Fall von bekannten λ und β .

Doch nicht nur die Schätzung der Lokations- und Dispersionseffekte, sondern auch ihre Identifikation ist in nichtwiederholten Designs eine große Herausforderung. In Kapitel 4.2 wird ein neues iteratives Verfahren, welches auf den Devianzdifferenzen beruht, vorgeschlagen. Dabei ist das iterative Vorgehen so aufgebaut, dass es die möglichen Vermischungen zwischen Dispersions- und Lokationseffekten berücksichtigt. Das vorgeschlagene Vorgehen weist einige Vorteile gegenüber dem von Pan und Taam (2002) auf. Es beruht nur auf einer einzigen Methode, welche nicht auf die Normalverteilungsannahme beschränkt ist, das Vorgehen ist klarer beschrieben und hat ein viel geringeres Risiko, mit zu großen

Modellen arbeiten zu müssen. Die Simulationsergebnisse zeigen dabei, dass jeder Schritt des neuen Verfahrens zumindest in einigen der betrachteten Situationen eine deutliche Sensitivitätssteigerung bei akzeptabler Erhöhung der Fehlerrate hervorruft. Für die Identifikation der Dispersionseffekte schneidet das vorgeschlagene Verfahren im Vergleich zu dem von Pan und Taam ähnlich gut ab. Der Grund für die geringere Sensitivität für die Identifikation der aktiven Lokationseffekte ist nicht klar.

Das auf den Devianzdifferenzen basierende iterative Verfahren zur Identifikation von aktiven Lokations- und Dispersionseffekten ist noch in mehreren Hinsichten weiter zu untersuchen. Zum Einen sollte überprüft werden, ob Schritt 1 nicht doch durch eine Identifikation basierend auf dem ASE durchgeführt werden sollte. Zum Anderen ist die Identifizierung aktiver Dispersionseffekte in Schritt 2 mit dem Verfahren von Bergman und Hynén (1997) zu vergleichen.

Ein großer Vorteil dieses Verfahrens ist die Unabhängigkeit von der Normalverteilung. Somit wäre zu untersuchen, wie gut es bei Unterstellung von anderen Verteilungen ist, und wie dann der Vergleich mit den Standardverfahren zur Identifikation aktiver Lokations- oder Dispersionseffekte ausfällt. Dabei kann das vorgeschlagene Verfahren dann auf die zur EQL, zur PL oder einer ihrer restringierten Varianten basierenden Devianzen übertragen werden.

Ein weiterer Punkt wäre zu untersuchen, ob gleichzeitig mit der Effektidentifikation auch eine Schätzung des Transformationsparameters λ erfolgen kann. Dabei wäre es sinnvoll, λ mit maximal einem Dispersionseffekt zu schätzen, siehe Kunert und Lehmkuhl (1998), und dann für festes λ auf weitere Dispersionseffekte zu testen. Dabei ist jedoch zu bedenken,

dass auch hier wieder eine Vermischung bestehen kann. Hat ein Dispersionseffekt sowohl einen positiven Einfluss auf den Erwartungswert als auch auf die Varianz, dann kann der Dispersionseffekt fälschlicherweise auch als eine vom Erwartungswert abhängige Varianz aufgefasst werden.

Anhang A

Regularitätsbedingungen

A.1 Alternative Regularitätsbedingung

Zunächst werden im folgenden Lemma die zu Regularitätsbedingung (R4) alternativen Regularitätsbedingungen von Erdbrügge und Auer (2010) angegeben. Es stellt Bedingungen an die Form der dritten Ableitungen und den Stichprobenraum, unter denen die Regularitätsbedingung (R4) erfüllt ist.

Lemma 2 (*Unbekannte Parameter λ und β*)

Es sei mit $\ell^(\mu, \phi | \mathbf{y})$ die Log-Likelihood $\ell(\mu, \phi | \mathbf{y})$ oder eine ihrer Erweiterungen bezeichnet. Erwartungswert und Varianz der Zielvariablen seien gegeben durch*

$$E\{y_i\} = \mu_i = T_\lambda^{-1}(\mathbf{x}_i^\top \alpha) = \begin{cases} (\mathbf{x}_i^\top \alpha \lambda + 1)^{1/\lambda} & : \lambda \neq 0 \\ \exp\{\mathbf{x}_i^\top \alpha\} & : \lambda = 0 \end{cases}$$

und $\text{Var}\{y_i\} = \phi_i \mu_i^{2\beta}$ mit $\phi_i > 0$.

Die stetige Abbildung $\mathbf{T} : \Theta \times \mathcal{X} \times \mathcal{Z} \rightarrow \Theta_{\mu, \phi, \lambda, \beta}$ sei gegeben durch $\mathbf{T}(\theta, \mathbf{x}_i, \mathbf{z}_i) = \mathbf{T}(\alpha, \gamma, \lambda, \beta, \mathbf{x}_i, \mathbf{z}_i) = (T_\lambda^{-1}(\mathbf{x}_i^\top \alpha), \exp\{\mathbf{z}_i^\top \gamma\}, \lambda, \beta)^\top = (\mu_i, \phi_i, \lambda, \beta)^\top$, erneut mit $\mu_i, \phi_i > 0$, wobei \mathcal{X} und \mathcal{Z} die Stichprobenräume der Lokations- und Dispersionseffekte sind. Diese Stichprobenräume seien beschränkt für $n \rightarrow \infty$, konkret gelte: $0 < \det[\frac{1}{n} \mathbf{X}^\top \mathbf{X}] < c_X < \infty$ und $0 < \det[\frac{1}{n} \mathbf{Z}^\top \mathbf{Z}] < c_Z < \infty$ für $n \rightarrow \infty$ und alle $\mathbf{X} \in \mathcal{X}$ und $\mathbf{Z} \in \mathcal{Z}$.

Die dritte Ableitung von $\ell^*(\mu, \phi | y_i)$ bezüglich $\theta_j, \theta_k, \theta_l, j, k, l = 1, \dots, s$, sei wie folgt darstellbar:

$$\begin{aligned} & \frac{\partial^3}{\partial \theta_j \partial \theta_k \partial \theta_l} \ell^*(\mu, \phi, \lambda, \beta | y_i) \\ &= (y_i - \mu_i)^2 A_{jkl}(\mu_i, \phi_i, \lambda, \beta) + (y_i - \mu_i) B_{jkl}(\mu_i, \phi_i, \lambda, \beta) + C_{jkl}(\mu_i, \phi_i, \lambda, \beta), \end{aligned}$$

wobei die Funktionen $A(\mu_i, \phi_i, \lambda, \beta)$, $B(\mu_i, \phi_i, \lambda, \beta)$ und $C(\mu_i, \phi_i, \lambda, \beta)$ auf jeder abgeschlossenen Menge $[\mu_*, \mu^*] \times [\phi_*, \phi^*] \times [\lambda_*, \lambda^*] \times [\beta_*, \beta^*] \subset \Theta_{\mu, \phi, \lambda, \beta}$ stetig sind. Dann ist die Regularitätsbedingung (R4) erfüllt.

Beweis: siehe Erdbrügge und Auer (2010)

A.2 Nachweis der Regularitätsbedingungen für PL^L

Für den nachfolgenden Nachweis der Regularitätsbedingungen (R1) bis (R4) von Satz 2 definiere $\vartheta = (\alpha^\top, \gamma^\top)^\top$ und

$$PL^L(\theta | y_i) := PL^L(\alpha, \gamma | y_i, \lambda, \beta) = -\frac{1}{2} \log(2\pi) + \mathbf{z}_i^\top \gamma + 2\beta \log(\mu_i) + \frac{(y_i - \mu_i)^2}{\varphi_i \mu_i^{2\beta} (1 - h_{ii})}.$$

Bei der Benutzung der restringierten Pseudo-Likelihood Version nach Lee und Nelder wird eine Normalverteilung unterstellt. Der zugehörige Erwartungswert, die unterstellte Varianz sowie die sich daraus ergebenden dritten und vierten zentralen Momente lauten wie folgt:

$$E(y_i) = \mu_i = T_\lambda^{-1}(\mathbf{x}_i^\top \alpha) = \begin{cases} (\mathbf{x}_i^\top \alpha \lambda + 1)^{1/\lambda}, & \lambda \neq 0 \\ \exp(\mathbf{x}_i^\top \alpha) & , \lambda = 0 \end{cases}$$

$$E\{(y_i - \mu_i)^2\} = \text{Var}(y_i) = \varphi_i V(\mu_i) = \varphi_i \mu_i^{2\beta} (1 - h_{ii})$$

$$E\{(y_i - \mu_i)^3\} = 0 \quad (\text{symmetrische Verteilung})$$

$$\begin{aligned} E\{(y_i - \mu_i)^4\} &= 1 * 3 * \text{Var}(y_i)^2 \\ &= 3\varphi_i^2 \mu_i^{4\beta} (1 - h_{ii})^2. \end{aligned}$$

Nun erfolgt der Nachweis der Regularitätsbedingungen (R1)-(R4) für die PL^L für den Fall von bekannten λ und β . Zum Nachweis von (R4) wird die im vorigen Abschnitt A.1 angegebene alternative Regularitätsbedingung nachgewiesen.

(R1) Die PL^L kann als Log-Likelihood einer zugehörigen Normalverteilungsdichte interpretiert werden. Diese Dichtefunktion existiert und ist dreimal stetig differenzierbar bezüglich aller interessierender Parameter, wenn die PL^L existiert und differenzierbar ist, da

$$\frac{\partial}{\partial \vartheta} \exp(PL^L(\vartheta | y_i)) = \exp(PL^L(\vartheta | y_i)) \frac{\partial PL^L(\vartheta | y_i)}{\partial \vartheta}.$$

Diese Argumentation überträgt sich entsprechend auf höhere Ableitungen.

(R2) Zum Nachweis von (R2) werden die ersten und zweiten Ableitungen der PL^L benötigt. Dazu werden zunächst die Ableitungen von

$$\begin{aligned} h_{ii} &= (W^{1/2}DX(X^T DWDX)^{-1}X^T DW^{1/2})_{ii} \\ &= w_{ii}d_{ii}^2 \mathbf{x}_i^T (X^T DWDX)^{-1} \mathbf{x}_i \\ &= \frac{1}{\varphi_i \mu_i^{2\beta}} \mu_i^{2(1-\lambda)} \mathbf{x}_i^T (X^T DWDX)^{-1} \mathbf{x}_i \end{aligned}$$

bestimmt. Unter der Verwendung von Matrixdifferentiationsregeln, siehe Harville (1997) S. 289 ff., ergeben sich die folgenden Ableitungen:

$$\begin{aligned} \frac{\partial h_{ii}}{\partial \mu_i} &= 2(1 - \lambda - \beta) \frac{h_{ii}}{\mu_i} (1 - h_{ii}) \\ \frac{\partial h_{ii}}{\partial \varphi_i} &= \frac{1}{\varphi_i} h_{ii} (h_{ii} - 1) \\ \frac{\partial^2 h_{ii}}{\partial \mu_i^2} &= 2(1 - \lambda - \beta) \frac{h_{ii}}{\mu_i^2} (1 - h_{ii}) [2(1 - \lambda - \beta)(1 - 2h_{ii}) - 1] \\ \frac{\partial^2 h_{ii}}{\partial \mu_i \partial \varphi_i} &= 2(1 - \lambda - \beta) \frac{h_{ii}}{\varphi_i \mu_i} (h_{ii} - 1)(1 - 2h_{ii}) \\ \frac{\partial^2 h_{ii}}{\partial \varphi_i^2} &= 2 \frac{1}{\varphi_i^2} h_{ii} (h_{ii} - 1)^2. \end{aligned}$$

Nun lassen sich die ersten Ableitungen der PL^L wie folgt darstellen:

$$\begin{aligned} & \frac{\partial}{\partial \alpha_{p_1}} PL^L(\alpha, \gamma | y_i) \\ &= \left\{ \frac{1}{2}(1 - h_{ii})^{-1} \frac{\partial h_{ii}}{\partial \mu_i} - \beta \mu_i^{-1} + \frac{(y_i - \mu_i)}{\varphi_i \mu_i^{2\beta} (1 - h_{ii})} \right. \\ & \quad \left. + \frac{1}{2} \frac{(y_i - \mu_i)^2}{\varphi_i \mu_i^{2\beta} (1 - h_{ii})^2} \left[2\beta \mu_i^{-1} (1 - h_{ii}) - \frac{\partial h_{ii}}{\partial \mu_i} \right] \right\} \mu_i^{1-\lambda} x_{ip_1} \end{aligned}$$

$$\begin{aligned} & \frac{\partial}{\partial \gamma_{q_1}} PL^L(\alpha, \gamma | y_i) \\ &= \left\{ \frac{1}{2}(1 - h_{ii})^{-1} \frac{\partial h_{ii}}{\partial \varphi_i} - \frac{1}{2\varphi_i} \right. \\ & \quad \left. - \frac{1}{2} \frac{(y_i - \mu_i)^2}{\varphi_i \mu_i^{2\beta} (1 - h_{ii})^2} \left[-(1 - h_{ii}) + \varphi_i \frac{\partial h_{ii}}{\partial \varphi_i} \right] \right\} \varphi_i z_{iq_1} \end{aligned}$$

$$\begin{aligned} & \frac{\partial}{\partial \lambda} PL^L(\alpha, \gamma, \beta, | y_i) \\ &= \frac{1}{2}(1 - h_{ii})^{-1} \frac{\partial h_{ii}}{\partial \lambda} - \frac{\beta}{\lambda} \left(\frac{x_i^\top \alpha}{\mu_i^\lambda} - \log \mu_i \right) + \frac{(y_i - \mu_i) \mu_i}{\varphi_i \mu_i^{2\beta} (1 - h_{ii})} \frac{1}{\lambda} \left(\frac{x_i^\top \alpha}{\mu_i^\lambda} - \log \mu_i \right) \\ & \quad + \frac{1}{2} \frac{(y_i - \mu_i)^2}{\varphi_i \mu_i^{2\beta} (1 - h_{ii})^2} \left[2 \frac{\beta}{\lambda} \left(\frac{x_i^\top \alpha}{\mu_i^\lambda} - \log \mu_i \right) (1 - h_{ii}) - \frac{\partial h_{ii}}{\partial \lambda} \right]. \end{aligned}$$

Hiermit und mit den obigen zentralen Momenten lässt sich Teil a) von Regularitätsbedingung (R1) zeigen:

$$\begin{aligned} E \left\{ \frac{\partial}{\partial \alpha_{p_1}} PL^L(\alpha, \gamma | y_i) \right\} &= 0, \\ E \left\{ \frac{\partial}{\partial \gamma_{q_1}} PL^L(\alpha, \gamma | y_i) \right\} &= 0. \end{aligned}$$

Die zweiten Ableitungen ergeben sich zu

$$\begin{aligned}
& \frac{\partial^2}{\partial \alpha_{p_1} \partial \alpha_{p_2}} PL^L(\alpha, \gamma | y_i) \\
&= \left\{ \frac{1}{2} \frac{\mu_i^{1-\lambda}}{(1-h_{ii})^2} \left(\frac{\partial h_{ii}}{\partial \mu_i} \right)^2 + \frac{1}{2} \frac{\mu_i^{1-\lambda}}{1-h_{ii}} \frac{\partial^2 h_{ii}}{\partial \mu_i^2} + \frac{1}{2} (1-\lambda) \frac{\mu_i^{-\lambda}}{1-h_{ii}} \frac{\partial h_{ii}}{\partial \mu_i} \right. \\
&\quad + \lambda \beta \mu_i^{-\lambda-1} - \frac{\mu_i^{1-\lambda-2\beta}}{\varphi_i (1-h_{ii})^2} \\
&\quad + \frac{y_i - \mu_i}{\mu_i^{2\beta} \varphi_i (1-h_{ii})^2} \left[\mu_i^{-\lambda} (1-\lambda-4\beta) (1-h_{ii}) + 2\mu_i^{1-\lambda} \frac{\partial h_{ii}}{\partial \mu_i} \right] \\
&\quad + \frac{(y_i - \mu_i)^2}{\mu_i^{2\beta} \varphi_i (1-h_{ii})^2} \left[-\lambda \beta \mu_i^{-1-\lambda} (1-h_{ii}) - 2\beta^2 \mu_i^{-2\lambda} (1-h_{ii}) \right. \\
&\quad \left. - \frac{1}{2} (1-\lambda-4\beta) \mu_i^{-\lambda} \frac{\partial h_{ii}}{\partial \mu_i} + 2\mu_i^{1-\lambda} (1-h_{ii})^{-1} \left(\frac{\partial h_{ii}}{\partial \mu_i} \right)^2 - \frac{1}{2} \mu_i^{1-\lambda} \frac{\partial^2 h_{ii}}{\partial \mu_i^2} \right] \left. \right\} \\
&\quad \cdot \mu_i^{1-\lambda} x_{ip_1} x_{ip_2}
\end{aligned}$$

$$\begin{aligned}
& \frac{\partial^2}{\partial \alpha_{p_1} \partial \gamma_{q_1}} PL^L(\alpha, \gamma | y_i) \\
&= \left\{ \frac{1}{2} (1-h_{ii})^{-2} \frac{\partial h_{ii}}{\partial \varphi_i} \frac{\partial h_{ii}}{\partial \mu_i} + \frac{1}{2} (1-h_{ii})^{-1} \frac{\partial^2 h_{ii}}{\partial \mu_i \partial \varphi_i} \right. \\
&\quad - \frac{y_i - \mu_i}{\varphi_i \mu_i^{2\beta} (1-h_{ii})} \left[(1-h_{ii})^{-1} \frac{\partial h_{ii}}{\partial \varphi_i} - \varphi_i^{-1} \right] \\
&\quad + \frac{(y_i - \mu_i)^2}{\varphi_i \mu_i^{2\beta} (1-h_{ii})^2} \left[\beta \mu_i^{-1} \frac{\partial h_{ii}}{\partial \varphi_i} - \frac{1}{2} \frac{\partial^2 h_{ii}}{\partial \mu_i \partial \varphi_i} - \beta \mu_i^{-1} (1-h_{ii}) \right. \\
&\quad \left. + \frac{1}{2} \varphi_i^{-1} \frac{\partial h_{ii}}{\partial \mu_i} - (1-h_{ii})^{-1} \frac{\partial h_{ii}}{\partial \varphi_i} \frac{\partial h_{ii}}{\partial \mu_i} \right] \left. \right\} \mu_i^{1-\lambda} \varphi_i x_{ip_1} z_{iq_1}
\end{aligned}$$

$$\begin{aligned}
& \frac{\partial^2}{\partial \gamma_{q_1} \partial \gamma_{q_2}} PL^L(\alpha, \gamma | y_i) \\
&= \left\{ \frac{1}{2} (1-h_{ii})^{-2} \left(\frac{\partial h_{ii}}{\partial \varphi_i} \right)^2 + \frac{1}{2} (1-h_{ii})^{-1} \frac{\partial^2 h_{ii}}{\partial \varphi_i^2} + \frac{1}{2} \varphi_i^{-2} \right. \\
&\quad - \frac{1}{2} \frac{(y_i - \mu_i)^2}{\mu_i^{2\beta}} \left[2\varphi_i^{-3} (1-h_{ii})^{-1} - 2\varphi_i^{-2} (1-h_{ii})^{-2} \frac{\partial h_{ii}}{\partial \varphi_i} \right. \\
&\quad \left. + 2\varphi_i^{-1} (1-h_{ii})^{-3} \left(\frac{\partial h_{ii}}{\partial \varphi_i} \right)^2 + \varphi_i^{-1} (1-h_{ii})^{-2} \frac{\partial^2 h_{ii}}{\partial \varphi_i^2} \right] \left. \right\} \varphi_i^2 z_{iq_1} z_{iq_2}
\end{aligned}$$

Somit ist auch Teil b) von Regularitätsbedingung (R2) erfüllt, denn mit den zentralen Momenten ergibt sich:

$$\begin{aligned}
& E \left\{ \frac{\partial}{\partial \alpha_{p_1}} PL^L(\alpha, \gamma | y_i) \cdot \frac{\partial}{\partial \alpha_{p_2}} PL^L(\alpha, \gamma | y_i) \right\} \\
&= E \left\{ -\frac{\partial^2}{\partial \alpha_{p_1} \partial \alpha_{p_2}} PL^L(\alpha, \gamma | y_i) \right\} \\
&= \left[\frac{1}{2} \frac{1}{(1-h_{ii})^2} \left(\frac{\partial h_{ii}}{\partial \mu_i} \right)^2 - \frac{2\beta \mu_i^{-1}}{1-h_{ii}} \frac{\partial h_{ii}}{\partial \mu_i} + 2\beta^2 \mu_i^{-2} + \frac{1}{\varphi_i \mu_i^{2\beta} (1-h_{ii})} \right] \mu_i^{2-2\lambda} x_{ip_1} x_{ip_2} \\
&= \left[2(1-\lambda-\beta)^2 h_{ii}^2 - 4(1-\lambda-\beta)\beta h_{ii} + 2\beta^2 + \frac{\mu_i^2}{\varphi_i \mu_i^{2\beta} (1-h_{ii})} \right] \mu_i^{-2\lambda} x_{ip_1} x_{ip_2}
\end{aligned}$$

$$\begin{aligned}
& E \left\{ \frac{\partial}{\partial \alpha_{p_1}} PL^L(\alpha, \gamma | y_i) \cdot \frac{\partial}{\partial \gamma_{q_1}} PL^L(\alpha, \gamma | y_i) \right\} \\
&= E \left\{ -\frac{\partial^2}{\partial \alpha_{p_1} \partial \gamma_{q_1}} PL^L(\alpha, \gamma | y_i) \right\} \\
&= \left[\frac{1}{2} \frac{1}{(1-h_{ii})^2} \frac{\partial h_{ii}}{\partial \mu_i} \frac{\partial h_{ii}}{\partial \varphi_i} - \frac{\beta \mu_i^{-1}}{1-h_{ii}} \frac{\partial h_{ii}}{\partial \varphi_i} - \frac{1}{2} \frac{1}{\varphi_i (1-h_{ii})} \frac{\partial h_{ii}}{\partial \mu_i} + \frac{\beta}{\mu_i \varphi_i} \right] \mu_i^{1-\lambda} \varphi_i x_{ip_1} z_{iq_1} \\
&= [-(1-\lambda-\beta)h_{ii}^2 - (1-\lambda-2\beta)h_{ii} + \beta] \mu_i^{-\lambda} x_{ip_1} z_{iq_1}
\end{aligned}$$

$$\begin{aligned}
& E \left\{ \frac{\partial}{\partial \gamma_{q_1}} PL^L(\alpha, \gamma | y_i) \cdot \frac{\partial}{\partial \gamma_{q_2}} PL^L(\alpha, \gamma | y_i) \right\} \\
&= E \left\{ -\frac{\partial^2}{\partial \gamma_{q_1} \partial \gamma_{q_2}} PL^L(\alpha, \gamma | y_i) \right\} \\
&= \left\{ \frac{1}{2} (1-h_{ii})^{-2} \left(\frac{\partial h_{ii}}{\partial \varphi_i} \right)^2 - \varphi_i^{-1} (1-h_{ii})^{-1} \frac{\partial h_{ii}}{\partial \varphi_i} + \frac{1}{2} \varphi_i^{-2} \right\} \varphi_i^2 z_{iq_1} z_{iq_2} \\
&= \left(\frac{1}{2} h_{ii}^2 + h_{ii} + \frac{1}{2} \right) z_{iq_1} z_{iq_2}
\end{aligned}$$

Die Grenzwerte der Ableitungen und ihrer Erwartungswerte für $\lambda \rightarrow 0$ lassen sich

durch folgende Grenzwertbetrachtungen, siehe Anhang B, bestimmen:

$$\lim_{\lambda \rightarrow 0} \frac{1}{\lambda} \left(\frac{x_i^\top \alpha}{\mu_i^\lambda} - \log \mu_i \right) = -\frac{1}{2} (x_i^\top \alpha)^2$$

$$\lim_{\lambda \rightarrow 0} \left(\frac{2}{\lambda^2} \left(\frac{\eta_i}{\mu_i^\lambda} - \log \mu_i \right) + \frac{1}{\lambda} \left(\frac{x_i^\top \alpha}{\mu_i^\lambda} \right)^2 \right) = -\frac{2}{3} (x_i^\top \alpha)^3.$$

(R3) Es resultiert folgende Informationsmatrix

$$\mathcal{I}(\alpha, \gamma, \lambda, \beta) = -E \begin{pmatrix} \frac{\partial^2 PL^L(\alpha, \gamma | y_i)}{\partial \alpha \partial \alpha} & \frac{\partial^2 PL^L(\alpha, \gamma | y_i)}{\partial \alpha \partial \gamma} \\ \frac{\partial^2 PL^L(\alpha, \gamma | y_i)}{\partial \gamma \partial \alpha} & \frac{\partial^2 PL^L(\alpha, \gamma | y_i)}{\partial \gamma \partial \gamma} \end{pmatrix}.$$

$$E \left(\frac{\partial^2 PL^L(\alpha, \gamma | y_i)}{\partial \alpha \partial \alpha} \right) = X^\top M_1 [M_2^L + 2\beta^2 I_n + 2(1 - \lambda - \beta)^2 H_1^2 - 4(1 - \lambda - \beta)\beta H_1] M_1 X$$

$$\begin{aligned} E \left(\frac{\partial^2 PL^L(\alpha, \gamma | y_i)}{\partial \alpha \partial \gamma} \right) &= E \left(\frac{\partial^2 PL_i^L}{\partial \gamma \partial \alpha} \right)^\top \\ &= X^\top M_1 [\beta I_n - (1 - \lambda - \beta)H_1^2 - (1 - \lambda - 2\beta)H_1] Z \end{aligned}$$

$$E \left(\frac{\partial^2 PL^L(\alpha, \gamma | y_i)}{\partial \gamma \partial \gamma} \right) = Z^\top \left(\frac{1}{2} H_1^2 + H_1 + \frac{1}{2} I_n \right) Z$$

mit den Matrizen $M_1 = \text{diag}(\mu_i^{-\lambda})$ und $M_2^L = \text{diag} \left(\frac{\mu_i^2}{\varphi_i \mu_i^{2\beta} (1 - h_{ii})} \right)$ sowie $H_1 = \text{diag}(h_{ii})$.

Fall 1: $\lambda = 1 - \beta$ bekannt

Im Fall, dass $\lambda = 1 - \beta$ bekannt reduziert sich die Informationsmatrix \mathcal{I}^L zu folgender

Matrix

$$\begin{aligned}
& \mathcal{I}_1^L(\alpha, \gamma) \\
&= \begin{pmatrix} \mathbf{X}^\top M_1 [M_2^L + 2\beta^2 I_n] M_1 \mathbf{X} & \beta \mathbf{X}^\top M_1 [I_n + H_1] \mathbf{Z} \\ \beta \mathbf{Z}^\top [I_n + H_1] M_1 \mathbf{X} & \frac{1}{2} \mathbf{Z}^\top [H_1^2 + 2H_1 + I_n] \mathbf{Z} \end{pmatrix} \\
&= \begin{pmatrix} \mathcal{I}_\alpha & \mathcal{I}_{\alpha\gamma} \\ \mathcal{I}_{\gamma\alpha} & \mathcal{I}_\gamma \end{pmatrix}.
\end{aligned}$$

Laut Harville (1997), S.99, existiert die Inverse $\mathcal{I}^{-1}(\alpha, \gamma)$, wenn die Matrix $Q := \mathcal{I}_\alpha - \mathcal{I}_{\alpha,\gamma} \mathcal{I}_\gamma^{-1} \mathcal{I}_{\gamma,\alpha}$ nichtsingulär ist. Für Q ergibt sich:

$$\begin{aligned}
Q &= \mathbf{X}^\top M_1 [M_2^L + 2\beta^2 I_n] M_1 \mathbf{X} \\
&\quad - \beta \mathbf{X}^\top M_1 [I_n + H_1] \mathbf{Z} \left[\frac{1}{2} \mathbf{Z}^\top (H_1^2 + 2H_1 + I_n) \mathbf{Z} \right]^{-1} \beta \mathbf{Z}^\top [I_n + H_1] M_1 \mathbf{X} \\
&= \mathbf{X}^\top M_1 [M_2^L + 2\beta^2 I_n] M_1 \mathbf{X} \\
&\quad - 2\beta^2 \mathbf{X}^\top M_1 (H_1 + I_n) \mathbf{Z} \left[\mathbf{Z}^\top (H_1 + I_n)^\top (H_1 + I_n) \mathbf{Z} \right]^{-1} \mathbf{Z}^\top (H_1 + I_n)^\top M_1 \mathbf{X} \\
&= \mathbf{X}^\top M_1 [M_2^L + 2\beta^2 I_n] M_1 \mathbf{X} - 2\beta^2 \mathbf{X}^\top M_1 \tilde{H}_Z M_1 \mathbf{X} \\
&= \mathbf{X}^\top M_1 [M_2^L + 2\beta^2 (I_n - \tilde{H}_Z)] M_1 \mathbf{X},
\end{aligned}$$

wobei \tilde{H}_Z die Hat-Matrix von $(H_1 + I_n)Z$ ist, vorausgesetzt $(H_1 + I_n)Z$ besitzt vollen Spaltenrang. Nun ist M_2^L als Diagonalmatrix mit lauter positiven Einträgen positiv definit und $(I_n - \tilde{H}_Z)$ als orthogonaler Projektor nichtnegativ definit. Somit ist ihre Summe positiv definit. Besitzt nun $\mathbf{X}^\top M_1$ vollen Spaltenrang, so ist Q nach Harville (1997), S. 213, positiv definit.

Fall 2: λ, β bekannt und unabhängig

Sind λ und β bekannt, aber gilt nicht zwingend, dass $\lambda = 1 - \beta$, so lautet die

Informationsmatrix wie folgt:

$$\begin{aligned} & \mathcal{I}_2^L(\alpha, \gamma) \\ &= \begin{pmatrix} \mathcal{I}_\alpha & \mathcal{I}_{\alpha\gamma} \\ \mathcal{I}_{\gamma\alpha} & \mathcal{I}_\gamma \end{pmatrix}. \end{aligned}$$

Dabei ist

$$\begin{aligned} \mathcal{I}_\alpha &= \mathbf{X}^\top M_1 \left[M_2^L + 2\beta^2 I_n + 2(1 - \lambda - \beta)^2 H_1^2 - 4(1 - \lambda - \beta)\beta H_1 \right] M_1 \mathbf{X} \\ \mathcal{I}_{\alpha\gamma} &= \mathbf{X}^\top M_1 \left[\beta I_n - (1 - \lambda - 2\beta)H_1 - (1 - \lambda - \beta)H_1^2 \right] \mathbf{Z} \\ \mathcal{I}_\gamma &= \frac{1}{2} \mathbf{Z}^\top (H_1^2 + 2H_1 + I_n) \mathbf{Z}. \end{aligned}$$

In diesem Fall ergibt sich die Matrix Q zu:

$$\begin{aligned} Q &= \mathbf{X}^\top M_1 \left[M_2^L + 2\beta^2 I_n + 2(1 - \lambda - \beta)^2 H_1^2 - 4(1 - \lambda - \beta)\beta H_1 \right] M_1 \mathbf{X} \\ &\quad - \mathbf{X}^\top M_1 \left[\beta I_n - (1 - \lambda - 2\beta)H_1 - (1 - \lambda - \beta)H_1^2 \right] \mathbf{Z} \left[\frac{1}{2} \mathbf{Z}^\top (H_1^2 + 2H_1 + I_n) \mathbf{Z} \right]^{-1} \\ &\quad \beta \mathbf{Z}^\top \left[\beta I_n - (1 - \lambda - 2\beta)H_1 - (1 - \lambda - \beta)H_1^2 \right] M_1 \mathbf{X} \\ &= \mathbf{X}^\top M_1 \left[M_2^L + 2\beta^2 I_n + 2(1 - \lambda - \beta)^2 H_1^2 - 4(1 - \lambda - \beta)\beta H_1 \right] M_1 \mathbf{X} \\ &\quad - 2\mathbf{X}^\top M_1 \left[\beta I_n - (1 - \lambda - \beta)H_1 \right] (H_1 + I_n) \mathbf{Z} \left[\mathbf{Z}^\top (H_1 + I_n)^\top (H_1 + I_n) \mathbf{Z} \right]^{-1} \\ &\quad \mathbf{Z}^\top (H_1 + I_n)^\top \left[\beta I_n - (1 - \lambda - \beta)H_1 \right] M_1 \mathbf{X} \\ &= \mathbf{X}^\top M_1 \left\{ M_2^L + 2 \left[\beta I_n - (1 - \lambda - \beta)H_1 \right]^2 \right\} M_1 \mathbf{X} \\ &\quad - 2\mathbf{X}^\top M_1 \left[\beta I_n - (1 - \lambda - \beta)H_1 \right] \tilde{H}_Z \left[\beta I_n - (1 - \lambda - \beta)H_1 \right] M_1 \mathbf{X} \\ &= \mathbf{X}^\top M_1 \left\{ M_2^L + 2 \left[\beta I_n - (1 - \lambda - \beta)H_1 \right] (I_n - \tilde{H}_Z) \left[\beta I_n - (1 - \lambda - \beta)H_1 \right] \right\} M_1 \mathbf{X} \end{aligned}$$

M_2^L ist positiv definit und der zweite Summand der eingeschachtelten Matrix ist nichtnegativ definit, da $(I_n - \tilde{H}_Z)$ als orthogonaler Projektor nichtnegativ definit ist

und die multiplizierenden Matrizen Vollrang besitzen, siehe Harville (1997) S. 216.

Somit ist Q , wenn M_1X Vollrang besitzt nichtsingulär.

(R4) Mit (R1) existieren alle dritten Ableitungen und es ist leicht zu sehen, dass alle 2.

Ableitungen in der folgenden Form dargestellt werden können:

$$\frac{\partial^2}{\partial\theta_j\partial\theta_k} PL^L = (y_i - \mu_i)^2 A_{jk}(\mu_i, \varphi_i, \lambda, \beta) + (y_i - \mu_i) B_{jk}(\mu_i, \varphi_i, \lambda, \beta) + C_{jk}(\mu_i, \varphi_i, \lambda, \beta).$$

Somit gilt für die dritten Ableitungen $\frac{\partial^3}{\partial\theta_j\partial\theta_k\partial\theta_l} PL^L$:

$$\begin{aligned} & \frac{\partial^3}{\partial\theta_j\partial\theta_k\partial\alpha_p} PL_i^L \\ &= \frac{\partial}{\partial\mu_i} \left(\frac{\partial^2}{\partial\theta_j\partial\theta_k} PL_i^L \right) \frac{\partial\mu_i}{\partial\eta_i} \frac{\partial\eta_i}{\partial\alpha_p} \\ &= \left[-2(y_i - \mu_i)\mu_i A_{jk}(\mu_i, \varphi_i, \lambda, \beta) + (y_i - \mu_i)^2 \frac{\partial}{\partial\mu_i} A_{jk}(\mu_i, \varphi_i, \lambda, \beta) \right. \\ & \quad \left. - B_{jk}(\mu_i, \varphi_i, \lambda, \beta) + (y_i - \mu_i) \frac{\partial}{\partial\mu_i} B_{jk}(\mu_i, \varphi_i, \lambda, \beta) \right. \\ & \quad \left. + \frac{\partial}{\partial\mu_i} C_{jk}(\mu_i, \varphi_i, \lambda, \beta) \right] \mu_i^{1-\lambda} x_{ip} \\ &= (y_i - \mu_i)^2 \underbrace{\frac{\partial}{\partial\mu_i} A_{jk}(\mu_i, \varphi_i, \lambda, \beta)}_{A_{jkl}(\mu_i, \varphi_i, \lambda, \beta)} \mu_i^{1-\lambda} x_{ip} \\ & \quad + (y_i - \mu_i) \underbrace{\left(\frac{\partial}{\partial\mu_i} B_{jk}(\mu_i, \varphi_i, \lambda, \beta) - 2\mu_i A_{jk}(\mu_i, \varphi_i, \lambda, \beta) \right)}_{B_{jkl}(\mu_i, \varphi_i, \lambda, \beta)} \mu_i^{1-\lambda} x_{ip} \\ & \quad + \underbrace{\left(\frac{\partial}{\partial\mu_i} C_{jk}(\mu_i, \varphi_i, \lambda, \beta) - B_{jk}(\mu_i, \varphi_i, \lambda, \beta) \right)}_{C_{jkl}(\mu_i, \varphi_i, \lambda, \beta)} \mu_i^{1-\lambda} x_{ip} \\ &= (y_i - \mu_i)^2 A_{jkl}(\mu_i, \varphi_i, \lambda, \beta) + (y_i - \mu_i) B_{jkl}(\mu_i, \varphi_i, \lambda, \beta) + C_{jkl}(\mu_i, \varphi_i, \lambda, \beta) \end{aligned}$$

$$\begin{aligned}
& \frac{\partial^3}{\partial\theta_j\partial\theta_k\partial\gamma_q} PL_i^L \\
&= \frac{\partial}{\partial\varphi_i} \left(\frac{\partial^2}{\partial\theta_j\partial\theta_k} PL_i^L \right) \frac{\partial\varphi_i}{\partial\xi_i} \frac{\partial\xi_i}{\partial\gamma_q} \\
&= (y_i - \mu_i)^2 \underbrace{\frac{\partial}{\partial\varphi_i} A_{jk}(\mu_i, \varphi_i, \lambda, \beta)}_{A_{jkl}(\mu_i, \varphi_i, \lambda, \beta)} \varphi_i z_{iq} \\
&\quad + (y_i - \mu_i) \underbrace{\frac{\partial}{\partial\gamma_q} B_{jk}(\mu_i, \varphi_i, \lambda, \beta)}_{B_{jkl}(\mu_i, \varphi_i, \lambda, \beta)} \varphi_i z_{iq} \\
&\quad + \underbrace{\frac{\partial}{\partial\varphi_i} C_{jk}(\mu_i, \varphi_i, \lambda, \beta)}_{C_{jkl}(\mu_i, \varphi_i, \lambda, \beta)} \varphi_i z_{iq} \\
&= (y_i - \mu_i)^2 A_{jkl}(\mu_i, \varphi_i, \lambda, \beta) + (y_i - \mu_i) B_{jkl}(\mu_i, \varphi_i, \lambda, \beta) + C_{jkl}(\mu_i, \varphi_i, \lambda, \beta).
\end{aligned}$$

Anhang B

Grenzwertbestimmungen für $\lambda \rightarrow 0$

Im Folgenden werden die im Nachweis der Regularitätsbedingungen nötigen Grenzwerte für $\lambda \rightarrow 0$ bestimmt.

Behauptung 8 Mit $\eta_i = x_i^\top \alpha$ und $\mu_i \neq 0$ gilt:

$$\lim_{\lambda \rightarrow 0} \frac{1}{\lambda} \left(\frac{\eta_i}{\mu_i^\lambda} - \log \mu_i \right) = -\frac{1}{2} \eta_i^2 \quad (\text{B.1})$$

Beweis:

Ohne Beschränkung der Allgemeinheit gilt $\lambda \neq 0$. Also ist folgender Grenzwert zu betrachten:

$$\lim_{\lambda \rightarrow 0} \frac{1}{\lambda} \left(\frac{\eta_i}{\eta_i \lambda + 1} - \frac{1}{\lambda} \log(\eta_i \lambda + 1) \right)$$

1. Fall: $\eta_i = 0$

$$\lim_{\lambda \rightarrow 0} \frac{1}{\lambda} \left(\frac{0}{0 \cdot \lambda + 1} - \frac{1}{\lambda} \log(0 \cdot \lambda + 1) \right) = \lim_{\lambda \rightarrow 0} \frac{1}{\lambda} (0 - 0) = 0 - 0 = \frac{1}{2} \eta_i^2$$

2. Fall: $\eta_i v \neq 0$

Es werden folgende Reihenentwicklung aus Bronstein (1993) benutzt:

$$(i) \log(1+x) = \sum_{k=1}^{\infty} (-1)^{k+1} \frac{x^k}{k} \text{ mit } |x| < 1$$

$$(ii) \frac{1}{a \pm x} = \sum_{k=0}^{\infty} \frac{(\mp x)^k}{a^{k+1}}$$

Damit ist

$$\begin{aligned} & \lim_{\lambda \rightarrow 0} \frac{1}{\lambda} \left(\frac{\eta_i}{\eta_i \lambda + 1} - \frac{1}{\lambda} \log(\eta_i \lambda + 1) \right) \\ &= \lim_{\lambda \rightarrow 0} \frac{1}{\lambda} \left(\sum_{k=0}^{\infty} (-1)^k \lambda^k \eta_i^{k+1} - \frac{1}{\lambda} \sum_{k=1}^{\infty} (-1)^{k+1} \eta_i^k \lambda^k \frac{1}{k} \right) \\ &= \lim_{\lambda \rightarrow 0} \frac{1}{\lambda} \left(\sum_{k=1}^{\infty} (-1)^{k-1} \lambda^{k-1} \eta_i^k - \sum_{k=1}^{\infty} (-1)^{k+1} \eta_i^k \lambda^{k-1} \frac{1}{k} \right) \\ &= \lim_{\lambda \rightarrow 0} \frac{1}{\lambda} \left(\sum_{k=1}^{\infty} (-1)^{k+1} \lambda^{k-1} \eta_i^k \left(1 - \frac{1}{k} \right) \right) \\ &= \lim_{\lambda \rightarrow 0} \sum_{k=1}^{\infty} (-1)^{k+1} \lambda^{k-2} \eta_i^k \left(1 - \frac{1}{k} \right) \\ &= \lim_{\lambda \rightarrow 0} \left((-1)^2 \lambda^{-1} \eta_i^1 (1-1) + \sum_{k=2}^{\infty} (-1)^{k+1} \lambda^{k-2} \eta_i^k \left(1 - \frac{1}{k} \right) \right) \\ &= \lim_{\lambda \rightarrow 0} \sum_{k=0}^{\infty} (-1)^{k+1} \lambda^k \eta_i^{k+2} \left(1 - \frac{1}{k+2} \right) \\ &= \lim_{\lambda \rightarrow 0} \left((-1)^1 \lambda^0 \eta_i^2 \left(1 - \frac{1}{2} \right) + \sum_{k=1}^{\infty} (-1)^{k+1} \lambda^k \eta_i^{k+2} \left(1 - \frac{1}{k+2} \right) \right) \\ &= \lim_{\lambda \rightarrow 0} \left(-\frac{1}{2} \eta_i^2 + \sum_{k=1}^{\infty} (-1)^{k+1} \lambda^k \eta_i^{k+2} \left(1 - \frac{1}{k+2} \right) \right) \\ &= -\frac{1}{2} \eta_i^2 + \sum_{k=1}^{\infty} \lim_{\lambda \rightarrow 0} (-1)^{k+1} \lambda^k \eta_i^{k+2} \left(1 - \frac{1}{k+2} \right) \\ &= -\frac{1}{2} \eta_i^2 \end{aligned}$$

q.e.d.

Behauptung 9 Mit $\eta_i = x_i^\top \alpha$ und $\mu_i \neq 0$ gilt:

$$\lim_{\lambda \rightarrow 0} \left(\frac{2}{\lambda^2} \left(\frac{\eta_i}{\mu_i^\lambda} - \log \mu_i \right) + \frac{1}{\lambda} \left(\frac{\eta_i}{\mu_i^\lambda} \right)^2 \right) = -\frac{2}{3} \eta_i^3 \quad (\text{B.2})$$

Beweis:

Ohne Beschränkung der Allgemeinheit gilt $\lambda \neq 0$ und somit $\mu_i = (\eta_i \lambda + 1)^{\frac{1}{\lambda}}$

$$\begin{aligned} & \lim_{\lambda \rightarrow 0} \left(\frac{2}{\lambda^2} \left(\frac{\eta_i}{\mu_i^\lambda} - \log \mu_i \right) + \frac{1}{\lambda} \left(\frac{\eta_i}{\mu_i^\lambda} \right)^2 \right) \\ &= \lim_{\lambda \rightarrow 0} \left(\frac{2}{\lambda^2} \left(\frac{\eta_i}{\eta_i \lambda + 1} - \frac{1}{\lambda} \log(\eta_i \lambda + 1) \right) + \frac{1}{\lambda} \left(\frac{\eta_i}{\eta_i \lambda + 1} \right)^2 \right) \end{aligned}$$

1. Fall: $\eta_i = 0$

$$\begin{aligned} & \lim_{\lambda \rightarrow 0} \left(\frac{2}{\lambda^2} \left(\frac{\eta_i}{\eta_i \lambda + 1} - \frac{1}{\lambda} \log(\eta_i \lambda + 1) \right) + \frac{1}{\lambda} \left(\frac{\eta_i}{\eta_i \lambda + 1} \right)^2 \right) \\ &= \lim_{\lambda \rightarrow 0} \left(\frac{2}{\lambda^2} \left(\frac{0}{0 \cdot \lambda + 1} - \frac{1}{\lambda} \log(0 \cdot \lambda + 1) \right) + \frac{1}{\lambda} \left(\frac{0}{0 \cdot \lambda + 1} \right)^2 \right) \\ &= \lim_{\lambda \rightarrow 0} \left(\frac{2}{\lambda^2} (0 - 0) + \frac{1}{\lambda} \cdot 0 \right) \\ &= 0 = -\frac{2}{3} \eta_i^3 \end{aligned}$$

2. Fall: $\eta_i \neq 0$

Zusätzlich zu den Reihenentwicklungen (i) und (ii) aus dem Beweis von Behauptung B.1 wird

noch folgende Reihenentwicklung benötigt:

$$\begin{aligned}
\text{(iii)} \quad \left(\frac{\eta_i}{\eta_i \lambda + 1} \right)^2 &= \frac{\eta_i^2}{\eta_i^2 \lambda^2 + 2\eta_i \lambda + 1} \\
&= \frac{1}{\lambda^2 + \frac{2\lambda}{\eta_i} + \frac{1}{\eta_i^2}} \\
&= \left(\frac{1}{\eta_i^2} \right)^{-1} (\lambda^2 \eta_i^2 + 2\lambda \eta_i + 1)^{-1} \\
&= \left(\frac{1}{\eta_i^2} \right)^{-1} \sum_{k=0}^{\infty} (-\lambda^2 \eta_i^2 + 2\lambda \eta_i)^k \quad \text{siehe Bronstein (1993)} \\
&= \left(\frac{1}{\eta_i^2} \right)^{-1} \sum_{k=0}^{\infty} (-1)^k (\lambda \eta_i)^k (\lambda \eta_i + 2)^k \\
&= \sum_{k=0}^{\infty} (-1)^k \lambda^k \eta_i^{k+2} (\lambda \eta_i + 2)^k
\end{aligned}$$

Hiermit gilt:

$$\begin{aligned}
&\frac{2}{\lambda^2} \left(\frac{\eta_i}{\mu_i^\lambda} - \log \mu_i \right) + \frac{1}{\lambda} \left(\frac{\eta_i}{\mu_i^\lambda} \right)^2 \\
&= \frac{2}{\lambda^2} \left(\frac{\eta_i}{\eta_i \lambda + 1} - \frac{1}{\lambda} \log(\eta_i \lambda + 1) \right) + \frac{1}{\lambda} \left(\frac{\eta_i}{\eta_i \lambda + 1} \right)^2 \\
&= \frac{2}{\lambda^2} \left(\frac{1}{\lambda + \frac{1}{\eta_i}} - \frac{1}{\lambda} \log(\eta_i \lambda + 1) \right) + \frac{1}{\lambda} \left(\frac{\eta_i}{\eta_i \lambda + 1} \right)^2 \\
&= \frac{2}{\lambda^2} \left(\sum_{k=0}^{\infty} (-1)^k \lambda^k \eta_i^{k+1} - \frac{1}{\lambda} \sum_{k=1}^{\infty} (-1)^{k+1} \eta_i^k \lambda^k \frac{1}{k} \right) + \frac{1}{\lambda} \sum_{k=0}^{\infty} (-1)^k \lambda^k \eta_i^{k+2} (\lambda \eta_i + 2)^k \\
&= 2 \left(\sum_{k=0}^{\infty} (-1)^k \lambda^{k-2} \eta_i^{k+1} - \sum_{k=1}^{\infty} (-1)^{k+1} \eta_i^k \lambda^{k-3} \frac{1}{k} \right) + \sum_{k=0}^{\infty} (-1)^k \lambda^{k-1} \eta_i^{k+2} (\lambda \eta_i + 2)^k \\
&= 2 \left(\sum_{k=1}^{\infty} (-1)^{k+1} \lambda^{k-3} \eta_i^k - \sum_{k=1}^{\infty} (-1)^{k+1} \eta_i^k \lambda^{k-3} \frac{1}{k} \right) + \sum_{k=2}^{\infty} (-1)^k \lambda^{k-3} \eta_i^k (\lambda \eta_i + 2)^{k-2} \\
&= 2 \left(\sum_{k=1}^{\infty} (-1)^{k+1} \lambda^{k-3} \eta_i^k - \sum_{k=1}^{\infty} (-1)^{k+1} \lambda^{k-3} \eta_i^k \frac{1}{k} \right) - \sum_{k=1}^{\infty} (-1)^{k+1} \lambda^{k-3} \eta_i^k (\lambda \eta_i + 2)^{k-2} \\
&\quad + (-1)^2 \lambda^{-2} \eta_i^1 (\lambda \eta_i + 2)^{-1}
\end{aligned}$$

$$\begin{aligned}
&= \sum_{k=1}^{\infty} (-1)^{k+1} \lambda^{k-3} \eta_i^k \left(2 - \frac{2}{k} - (\lambda \eta_i + 2)^{k-2} \right) + \frac{\eta_i}{\lambda^2} (\lambda \eta_i + 2)^{-1} \\
&= \sum_{k=4}^{\infty} (-1)^{k+1} \lambda^{k-3} \eta_i^k \left(2 - \frac{2}{k} - (\lambda \eta_i + 2)^{k-2} \right) + \lambda^{-2} \eta_i (2 - 2 - (\lambda \eta_i + 2)^{-1}) \\
&\quad - \lambda^{-1} \eta_i^2 (2 - 1 - (\lambda \eta_i + 2)^0) + \lambda^0 \eta_i^3 \left(2 - \frac{2}{3} - (\lambda \eta_i + 2) \right) + \frac{\eta_i}{\lambda^2} (\lambda \eta_i + 2)^{-1} \\
&= \sum_{k=0}^{\infty} (-1)^{k+1} \lambda^{k+1} \eta_i^{k+4} \left(2 - \frac{2}{k+4} - (\lambda \eta_i + 2)^{k+2} \right) - \eta_i \lambda^{-2} (\lambda \eta_i + 2)^{-1} - 0 \\
&\quad + \eta_i^3 \left(-\frac{2}{3} - \lambda \eta_i \right) + \eta_i \lambda^{-2} (\lambda \eta_i + 2)^{-1} \\
&= \sum_{k=0}^{\infty} (-1)^{k+1} \lambda^{k+1} \eta_i^{k+4} \left(2 - \frac{2}{k+4} - (\lambda \eta_i + 2)^{k+2} \right) - \frac{2}{3} \eta_i^3 - \lambda \eta_i^4.
\end{aligned}$$

Somit gilt:

$$\begin{aligned}
&\lim_{\lambda \rightarrow 0} \left(\frac{2}{\lambda^2} \left(\frac{\eta_i}{\mu_i^\lambda} - \log \mu_i \right) + \frac{1}{\lambda} \left(\frac{\eta_i}{\mu_i^\lambda} \right)^2 \right) \\
&= \lim_{\lambda \rightarrow 0} \left(\sum_{k=0}^{\infty} (-1)^{k+1} \lambda^{k+1} \eta_i^{k+4} \left(2 - \frac{2}{k+4} - (\lambda \eta_i + 2)^{k+2} \right) - \frac{2}{3} \eta_i^3 - \lambda \eta_i^4 \right) \\
&= \sum_{k=0}^{\infty} \lim_{\lambda \rightarrow 0} (-1)^{k+1} \lambda^{k+1} \eta_i^{k+4} \left(2 - \frac{2}{k+4} - (\lambda \eta_i + 2)^{k+2} \right) - \frac{2}{3} \eta_i^3 - 0 \\
&= -\frac{2}{3} \eta_i^3 - 0
\end{aligned}$$

q.e.d.

Anhang C

Tabellen und Grafiken

C.1 Tabellen

Tabelle C.1: simulierte lokale Quantile für die Lokationseffekte (10000 Datensätze, 2^5 -Design)

Anzahl LE	Anzahl DE (Niveau)								
	0			1			2		
	0.8	0.9	0.95	0.8	0.9	0.95	0.8	0.9	0.95
0	1.7	2.9	4.1	2.0	3.3	4.6	2.1	3.5	5.1
1	1.8	3.1	4.5	2.0	3.3	4.7	2.2	3.6	5.3
2	1.8	3.0	4.4	2.1	3.5	4.9	2.4	3.9	5.4
3	1.8	3.1	4.5	2.2	3.6	5.3	2.6	4.3	6.0
4	1.8	3.2	4.7	2.5	4.1	5.7	2.9	4.7	6.7
5	1.8	3.3	4.8	2.7	4.6	6.5	3.4	5.7	7.9
6	1.9	3.4	5.0	2.7	4.4	6.1	3.3	5.5	7.6

Tabelle C.2: simulierte globale Quantile für die Lokationseffekte (2000 Datensätze, 2^4 -Design)

Anzahl LE	Anzahl DE (Niveau)								
	0			1			2		
	0.8	0.9	0.95	0.8	0.9	0.95	0.8	0.9	0.95
0	7.2	8.7	10.3	7.9	9.8	11.2	8.7	11.0	12.9
1	7.4	8.9	10.3	8.2	9.8	11.7	9.6	11.7	13.9
2	8.0	9.8	11.7	9.3	11.7	13.6	10.3	12.8	15.6
3	8.4	10.5	12.6	10.7	13.1	16.0	13.2	16.8	20.0
4	8.9	11.2	13.7	13.0	16.5	20.4	17.8	22.3	27.1
5	9.4	11.9	14.7	13.3	16.9	20.1	18.5	23.8	28.1
6	10.24	12.5	14.8	13.9	17.4	21.7	22.5	28.5	34.9

Tabelle C.3: simulierte globale Quantile für die Dispersionseffekte (2000 Datensätze, 2^4 -Design)

Anzahl LE	Anzahl DE (Niveau)								
	0			1			2		
	0.8	0.9	0.95	0.8	0.9	0.95	0.8	0.9	0.95
0	7.0	8.7	10.3	7.0	8.7	10.6	7.2	8.9	10.6
1	8.4	10.4	12.6	8.4	10.5	12.2	8.6	11.0	12.8
2	11.1	13.5	15.3	10.5	13.0	15.2	11.0	13.3	15.3
3	13.7	16.5	19.5	13.7	17.0	19.8	14.4	17.3	20.2
4	18.4	21.8	25.5	17.5	20.5	24.0	19.0	22.9	26.5
5	26.7	31.9	37.5	24.9	30.0	34.8	25.7	31.2	35.3
6	26.7	31.9	37.5	24.9	30.0	34.8	25.7	31.2	35.3

C.2 Grafiken

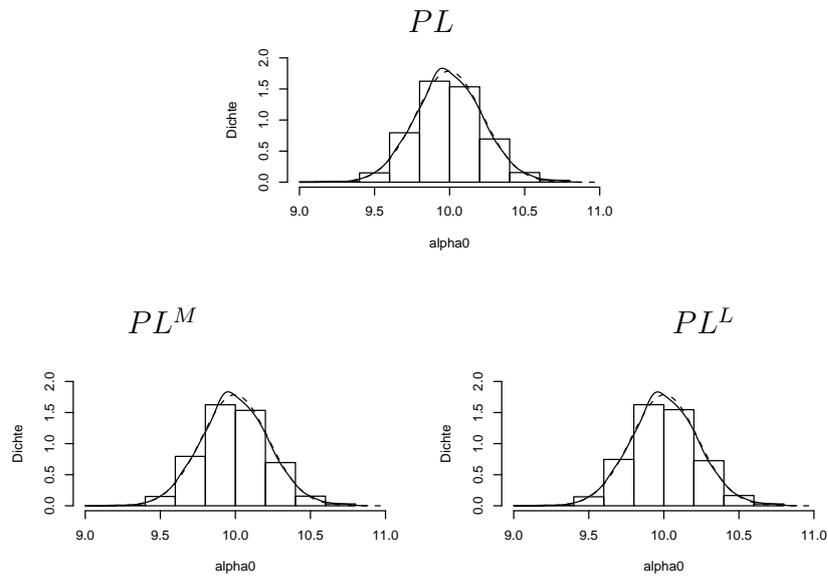


Abbildung C.1: Histogramm mit Dichteschätzung (—) und Normalverteilungsdichte (---) für α_0

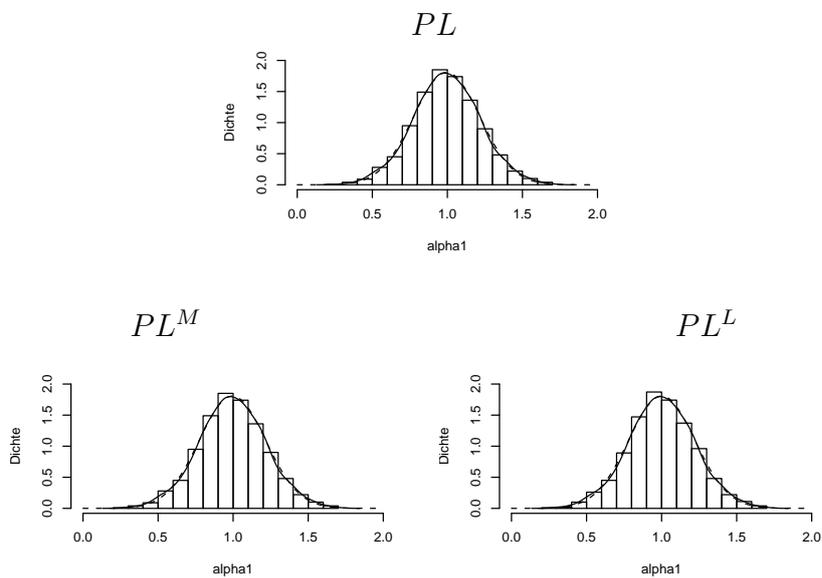


Abbildung C.2: Histogramm mit Dichteschätzung (—) und Normalverteilungsdichte (---) für α_1

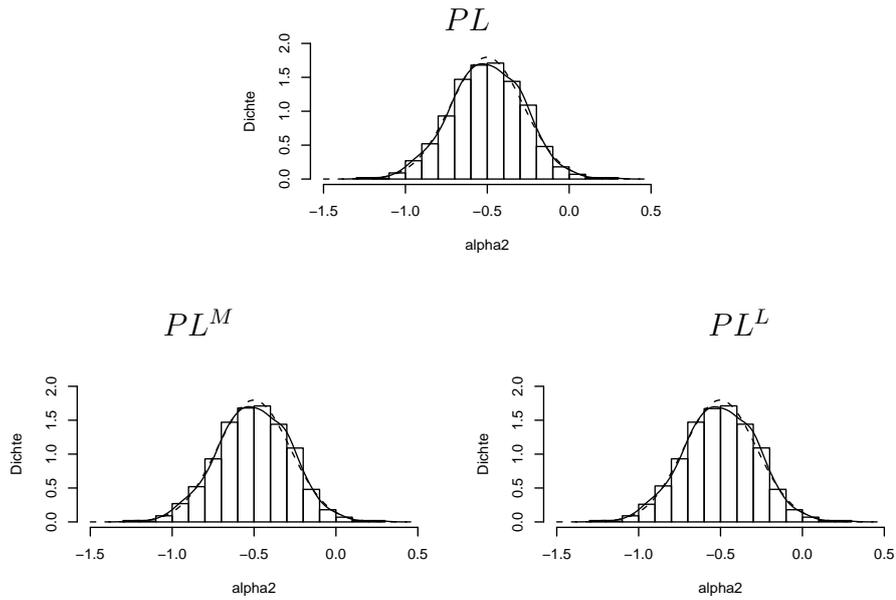


Abbildung C.3: Histogramm mit Dichteschätzung (—) und Normalverteilungsdichte (---) für α_2

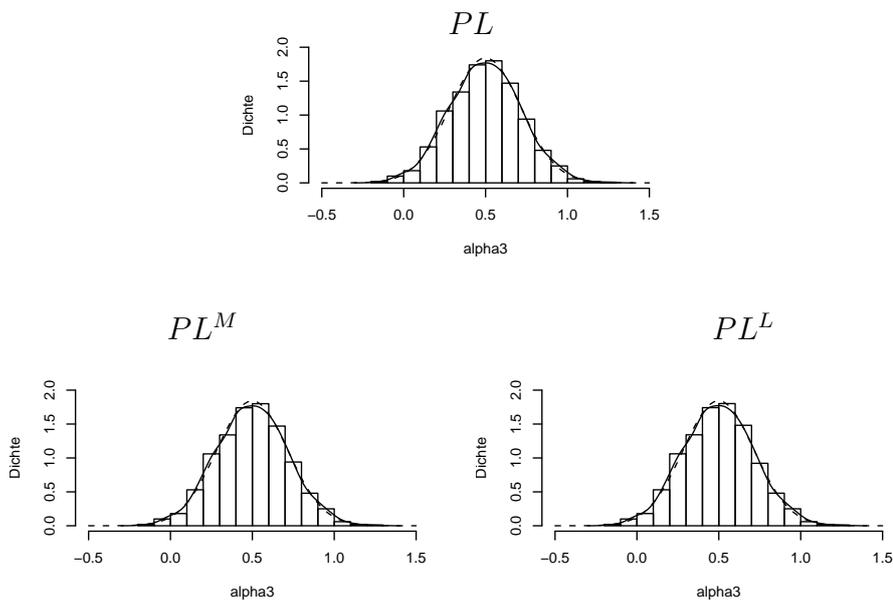


Abbildung C.4: Histogramm mit Dichteschätzung (—) und Normalverteilungsdichte (---) für α_3

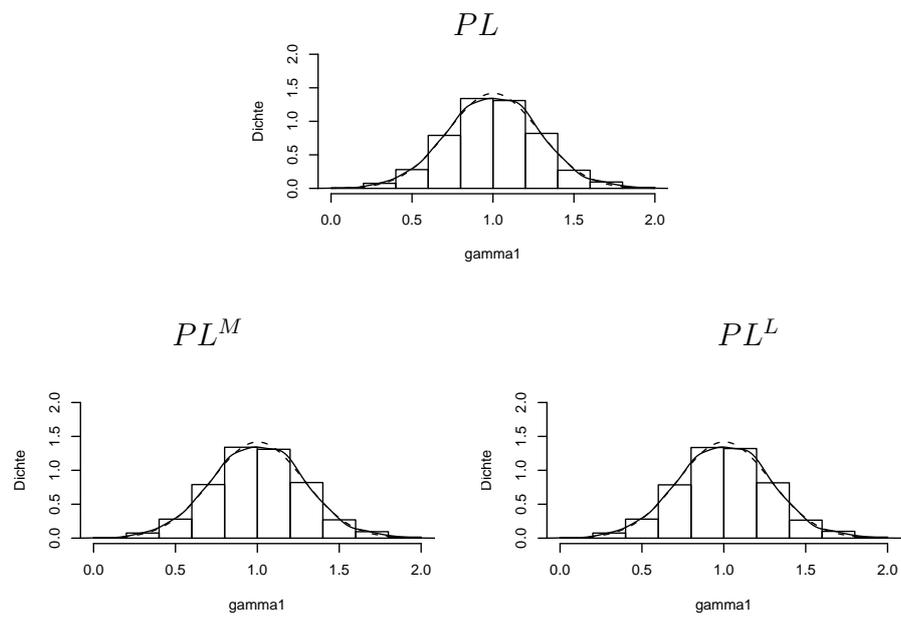


Abbildung C.5: Histogramm mit Dichteschätzung (—) und Normalverteilungsdichte (- - -)

für γ_1

Literatur

- Aitken, M. (1987). Modelling Variance Heterogeneity in Normal Regression using GLIM. *Applied Statistics* 36, 332–339.
- Auer, C., Erdbrügge, M. und Göbel, R. (2004). Comparison of Multivariate Methods for Robust Parameter Design in Sheet Metal Spinning. *Applied Stochastic Models in Business and Industry* 20, 201–218.
- Auer, C. und Kunert, J. (2006). On A Heuristic Analysis of Highly Fractionated 2^n Factorial Experiments. *Metrika* 63, 43–54.
- Bergman, B. und Hynén, A. (1997). Dispersion Effects from Unreplicated Designs in the 2^{k-p} Series. *Technometrics* 39, 191–198.
- Box, G.E.P. (1988). Signal-to-noise Ratios, Performance Criteria, and Transformations. *Technometrics* 30, 1–17.
- Box, G.E.P. und Cox, D.R. (1964). An Analysis of Transformations. *Journal of the Royal Statistical Society B* 26, 211–252.
- Bronstein, I.N., Semendjaev, K.A., Musiol, G. und Mühlig, H. (1993). *Taschenbuch der Mathematik*. Verlag Harri Deutsch, Thun, Frankfurt am Main.

- Cox, D. R. und Reid, N. (1987). Parameter Orthogonality and Approximate Conditional Inference. *Journal of the Royal Statistical Society B* 49, 1–39.
- Davidian, M. und Carroll, R. J. (1988). A Note on Extended Quasi-likelihood. *Journal of the Royal Statistical Society B* 50, 74–82.
- Dong, F. (1993). On the Identification of Active Contrasts in Unreplicated Fractional Factorials. *Statistica Sinica* 3, 209–217.
- Efron, B. (1986). Double exponential families and their use in generalized linear regression. *Journal of the American Statistical Association* 81, 709–721.
- Engel, J. und Huele, A. F. (1996). A Generalized Linear Modeling Approach to Robust Design. *Technometrics* 38, 365–373.
- Erdbrügge, M. (2003). *Verfahren zur Modellauswahl bei Experimenten mit Dispersionsfaktoren*. Logos Verlag, Berlin.
- Erdbrügge, M. und Auer, C. (2010). Generalized Linear Models and Transformations for Unreplicated Factorial Experiments in the Presence of Dispersion Effects. *Journal of Statistical Planning and Inference* 35, 2–12.
- Erdbrügge, M. (2002). Model Selection Strategies for Experiments with Dispersion Effects - Transformations vs. Generalized Linear Models. *SFB 475 Technical Report, Universität Dortmund 69/2002*.
- Fahrmeir, L. und Tutz, G. (1997). *Multivariate Statistical Modelling Based on Generalized Linear Models*, 3. Ausg. Springer.
- Firth, D. (1991). Chapter 3: Generalized linear models. In *Hinkley, Reid, Snell (Eds.):*

- Statistical Theory and Modelling. In honour of Sir David Cox, FRS*, pp. 55–82. Chapman & Hall.
- Fuller, H. T. und Bisgaard, S. (1995). A Comparison of Dispersion Effect Identification Methods for Unreplicated Two-Level Factorials. *University of Wisconsin-Madison, Center for Quality and Productivity Improvement Report 132*, 239–262.
- Harville, D. A. (1997). *Matrix Algebra From a Statistician's Perspective*. Springer.
- Kunert, J. (1997). On the Use of the Factor-Sparsity Assumption to Get an Estimate of the Variance in Saturated Designs. *Technometrics* 39, 81–90.
- Kunert, J., Auer, C., Erdbrügge, M. und Ewers, R. (2007). An Experiment to Compare Taguchi's Product Array and the Combined Array. *Journal of Quality Technology* 39, 17–34.
- Kunert, J. und Lehmkuhl, F. (1998). *MODA 5 - Advances in Model-Oriented Data Analysis and Experimental Design; The Generalized β -Method in Taguchi Experiments*, pp. 223–230. Atkinson, Pronzato, Wynn (Eds.), Physica-Verlag.
- Lee, Y. und Nelder, J. A. (1998). Generalized Linear Models for the Analysis of Quality-Improvement Experiments. *The Canadian Journal of Statistics* 26, 95–105.
- Lee, Y. und Nelder, J. A. (2000). The relationship between double-exponential families and extended quasi-likelihood families, with application to modelling Geissler's human sex ration data. *Journal of Applied Statistics* 49, 413–419.
- Lee, Y. und Nelder, J. A. (2003). Robust Design via Generalized Linear Models. *Journal of Quality Technology* 35, 2–12.

- Lehmann, E. L. (1983). *Theory of Point Estimation*. Wiley Series in Probability and Mathematical Statistics.
- Lehmann, E. L. (1986). *Testing Statistical Hypotheses*. Springer.
- Lenth, R. V. (1989). Quick and Easy Analysis of Unreplicated Factorials. *Technometrics* 31, 469–473.
- Logothetis, N. (1988). The Role of Data Transformation in Taguchi Analysis. *Quality and Reliability Engineering International* 4, 49–61.
- McCullagh, P. und Nelder, J. A. (1989). *Generalized Linear Models*. Chapman & Hall.
- McGrath, R. N. (2003). Separating Location and Dispersion Effects in Unreplicated Fractional Factorial Designs. *Journal of Quality Technology* 35, 306–316.
- McGrath, R. N. und Lin, D. K. J. (2002). A Nonparametric Dispersion Test for Unreplicated Two-Level Fractional Factorial Designs. *Nonparametric Statistics* 14, 699–714.
- McGrath, R. N. und Lin, D. K. J. (2001). Testing Multiple Dispersion Effects in Unreplicated Fractional Factorial Designs. *Technometrics* 43, 406–414.
- Nelder, J. A. (2000). Quasi-likelihood and Pseudo-likelihood Are not the Same Thing. *Journal of Applied Statistics* 27, 1007–1011.
- Nelder, J. A. und Lee, Y. (1991). Generalized Linear Models for the Analysis of Taguchi-type Experiments. *Applied Stochastic Models and Data Analysis* 7, 107–120.
- Nelder, J. A. und Lee, Y. (1992). Likelihood, Quasi-likelihood and Pseudolikelihood: Some Comparisons. *Journal of the Royal Statistical Society B* 54, 273–284.

- Nelder, J. A. und Pregibon, D. (1987). An Extended Quasi-likelihood Function. *Biometrika* 74, 221–232.
- Pan, G. und Taam, W. (2002). Confounding of Location and Dispersion Effects in Unreplicated Fractional Factorials. *Journal of Statistical Computation and Simulation* 72, 431–450.
- Raghavarao, D. und Altan, S. (2003). A heuristic analysis of highly fractionated 2^n factorial experiments. *Metrika* 58, 185–191.
- Richter, S. H., Garner, J. P., Auer, C., Kunert, J. und Würbel, H. (2010). Systematic variation improves reproducibility of animal experiments. *Nature Methods* 7, 167–168.
- Schervish, M. J. (1995). *Theory of Statistics*. Springer Series in Statistics.
- Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. John Wiley & Sons.
- Smyth, G. K., Huele, A. F. und Verbyla, A. P. (2001). Exact and Approximate REML for Heteroscedastic Regression. *Statistical Modelling* 1, 161–175.
- Steinberg, D. M. und Bursztyn, D. (1994). Dispersion Effects in Robust-design Experiments with Noise Factors. *Journal of Quality Technology* 26, 12–20.
- Toutenburg, H. (2003). *Lineare Modelle*. Physica-Verlag, Heidelberg, 2. Auflage.
- Toutenburg, H., Gössel, R. und Kunert, J. (1998). *Quality Engineering: Eine Einführung in Taguchi Methoden*. Prentice Hall, München.
- Wasan, M. T. (1970). *Parametric Estimation*. McGraw-Hill.

Danksagung

Hiermit möchte ich mich noch einmal ganz herzlich bei meinen Eltern und Schwiegereltern bedanken, die sich während meiner Promotionszeit immer ganz selbstverständlich und liebevoll um meine drei Kinder gekümmert haben, und bei meinem Mann, der mich immer wieder aufgebaut und an mich geglaubt hat. Ohne sie hätte ich die Promotion nie geschafft. Ein großer Dank geht auch an einen Chef, Herrn Kunert, der mich die ganzen Jahre über fachlich begleitet hat. Des Weiteren bedanke ich mich bei all meinen Kollegen, die immer für eine entspannte und freundliche Arbeitsatmosphäre gesorgt haben, so dass ich die Zeit an der Fakultät Statistik und insbesondere an diesem Lehrstuhl sehr genossen habe.