Seminararbeit:

Analye longitundinaler Daten unter Verwendung von generalisierten linearen Modellen

Autor: Markus Riedl

Matrikelnummer: 10600666

Betreuerin: Andrea Wienciez

May 30, 2014

Inhaltsverzeichnis

1 Einleitung

Die Analyse von longitudinalen Daten wird in vielen verschiedenen wissenschaftlichen Bereichen benötigt. Ein wichtiger Anwendungsbereich liegt zum Beispiel in verschiedenen klinischen Studien. Dort werden häufig Messungen an einem Patienten zu verschiedenen Zeitpunkten erhoben. Das Hauptproblem von Daten dieser Art ist, dass diese korreliert sind.
Zur Analyse solch einer Datenstruktur existierte bis zu der Erscheinung des Paper „Longitudinal Data Analysis Using Generalized Linear Models“, welches in Zusammenarbeit von Liang und Zeger im Jahre 1986 veröffentlicht wurde, keine allgemeingültige Methodologie.
In der folgenden Arbeit beziehe ich mich hauptsächlich auf dieses Werk und die dazugehörige Einleitung, welche von Peter J. Diggle verfasst wurde. Als Liang und Zeger an ihrer Abhandlung arbeiteten, stellte die Laird–Ware-Methodologie (1982) die essentiellen Anforderungen für die Analyse von longitudinalen Daten unter Normalverteilungsannahme bereit.
Auch für binäre Daten existierten bereits drei verschiedene Modelle mit Messwiederholungen. In diesen Modellen wird angenommen, dass die Beobachtungen von einem Subjekt austauschbare Korrelationen haben. Diese wurden von Ochi und Prentice (1984) unter Verwendung des Probit Link, von Stiratelli, Laird und Ware (1984) mittels Logit Link und von Koch et al. (1977) unter Verwendung von log linearen Modellen entwickelt. Wobei nur das Modell von Stiratelli, Laird und Ware zeitabhängige Kovariabeln zulässt.
Die Autoren Liang und Zeger waren jedoch von nicht-normalverteilten, korrelierten Daten betroffen für die derzeit noch keine Methodologie existierte. In der folgenden Arbeit wird erläutert, wie Liang und Zege das klassische generalisierte lineare Modell erweiterten, sodass dieses auch für korrelierte Daten gültig.
Wie essentiell der Beitrag dieser Abhandlung für die angewandte Statistik ist, wurde auch in meinem bisherigen Bachelor Studium deutlich. In der Vorlesung Generalisierte Regressionsmodelle von Prof. Dr. Küchenhoff wurde uns die Schätzung von longitudinalen mittels generalisierten linearen Modellen doziert. Auch in der Veranstaltung “statistisches Praktikum „lagen longitudinale Daten vor, was eine Model-

lierung mittels generalisierte lineare Modelle mit zufälligen Effekten erforderte.

2 Longitudinale Daten

Longitudinale Daten sind definiert, durch zeitlich wiederholte Messungen die an der selben experimentellen Einheiten erhoben wurden, welche im folgenden als Subjekt oder Individuen bezeichnet wird (Diggle, Liang und Zeger, 1986).
Datensätze dieser Art bestehen aus einer Zielvariable y_{it} , einen $p \times 1$ Vektor von Kovariablen x_{it}, die zu den Zeiten $t = 1, \ldots, n_i$ beobachtet wurden, für das Subjekt $i = 1, \ldots, K$.. In vielen Anwendungensbereichen ist es das primäre Ziel Rückschlüsse über die Mittelwerte, $\mu_{ij} = E(Y_{iy})$, zu ziehen, welche möglicherweise von der Zeit, der Behandlungszuteilung oder von einer Anzahl von anderen Kovariablen abhängen. Erklärende Variablen werden in diesem Kontext im allgemeinen zu individuellen Subjekten, individuellen Zeiten oder zu individuellen Reaktionen zugeordnet. Longitudinale Daten bieten im Vergleich zu Querschnittdaten den Vorteil, dass individuelle Änderungen über die Zeit erfasst werden können (Diggle, Liang und Zeger, 1986).
Im Gegensatz zu der klassischen Zeitreihenanalyse steht bei der Analyse von longitudinalen Daten die Schätzung von Kovariableneffekten im Vordergrund (Anne-Laure Boulesteix). Ein weiterer wichtiger Unterschied zwischender Anaylse von longitudinalen Daten und der klassischen Zeitreihenanalyse ist, dass sich letztere fast ausschließlich auf die Analyse von langen, einzelnen Reihen konzentriert, während typische longitudinale Datensätze aus vielen kurzen Datenreihen bestehen. Desweiteren ist in longitudinalen Studien nicht die Kovarianzstruktur der Daten von direktem Interesse, was in der klassischen Zeitreihenanalyse der Fall ist (Diggle, Liang und Zeger, 1986).

2.1 Analyse longitudinaler Daten

Das wissenschaftliche Interesse liegt meist in der Analyse von zeitlich veränderten Mustern oder in der Abhängigkeit von den Ergebnissen auf die Kovariaten (Diggle, Liang und Zeger, 1986).
Hierbei sind Messungen an dem selben Individuum typischerweise ähnlicher, als

Messungen an verschiedenen Subjekten. Deshalb besteht keine Unabhängigkeit der Daten mehr und die Korrelation der Werte von einem Individuum muss mit in die Berechnung aufgenommen werden (Diggle, Liang und Zeger, 1986).

Eine Schwierigkeit in der Analyse von nicht normalverteilten longitudinalen Daten, ist der Mangel von ergiebigen Klassen von Modellen (Diggle, Liang und Zeger, 1986). Liang und Zeger versuchten ein generalisiertes lineares Modell für die Randverteilung von y_{it} zu verwenden. Sie spezifizierten keine Form für die gemeinsame Verteilung der wiederholten Messungen. Stattdessen leiteten sie Schätzgleichungen her, die konsistente Schätzer für die Regressionsparameter und ihrer Varianz, unter schwachen Annahmen über die gemeinsame Verteilung liefern(Diggle, Liang und Zeger, 1986).

3 Generalisierte Lineare Modelle

Da Liang und Zeger mit Daten auseinandergesetzt waren, die nicht normalverteilt sind schien ihnen ein generalisierte lineare Modell geeignet zu erscheinen. Dieses Modell bietet den Vorteil, dass Beobachtungen nicht notwendigerweise normalverteilt sein müssen.
Ein generalisiertes oder auch verallgemeinertes linears Modell hat folgende Form und Eigenschaften.
Zum einem wird in einem generalisierten linearen Modell die Verteilungsannahme getroffen, dass die Dichte von y_i aus einer Exponentialfamilie stammt. Mögliche Verteilungen aus der Klasse der Exopnentialfamilie sind die Normal-, Binomial-, Multinomial-, Poisson-, oder Gammaverteilung. Durch die Strukturannahme wird der Erwartungswert $\mu_i = E(y_i)$ mit dem linearen Prädiktor $\eta_i = x_i'\beta$ durch eine invertierbaren Response-Funktion h mit

$$\mu_i = h(\eta_i) = h(x_i'\beta)$$

verbunden (Funk-Hüsges). Die inverse Response-Funktion $g = h^{-1}$ heißt Link-Funktion und ist definiert durch:

$$\eta_i = (g(\mu_i)).$$

(Prof. Dr. Küchenhoff). Die Dichtefunktion einer Exponentialfamilie ist wie folgt definiert:

$$f(y_i|\theta_i) = exp\left(\frac{y_i\theta_i - b(\theta_i)}{\phi} + c(y_i, \phi)\right),$$

Hierbei bezeichent ϕ den natürlichen oder kanonischen Parameter, $b(\theta)$ die Noramlisierung, ϕ den Skalen- oder Dispersionsparameter und $c(y_i, \phi)$ eine Funktion (Prof. Dr. Küchenhoff). Für Verteilungen aus der Exponentialfamilie gilt,

$$E(y_i) = b'(\theta_i)$$

$$Var(y_i) = \phi b''(\theta_i).$$

Der kanonische Parameter θ_i der Exponentialfamilie st abhängig von dem Erwartungswert $E(Y_i) = \mu_i$. Gilt die Beziehung

$$\theta_i = g(\mu_i),$$

dann ist g die natürliche Link-Funktion und es folgt:

$$\theta_i = x_i'\beta.$$

(Prof. Dr. Küchenhoff).
Auch Liang und Zeger nehmen für ihre folgenden Analysen, mit einer leicht abweichenden Notation an, dass die marginale Dichte von y_{it} gegeben ist durch:

$$f(y_{it}) = exp[\{y_{it}\theta_{it} - a(\theta_{it}) + b(y_{it})\}\phi], \tag{3.1}$$

wobei $\theta_{it} = h(\eta_{it})$ und $\eta_{it} = x_{it}\beta$.
Anhand dieser Ausarbeitung sind die ersten zwei Momente von y_{it} gegeben durch

$$E(y_{it}) = a'(\theta_{it}) \quad \text{und} \quad var(y_{it}) = \frac{a''(\theta_{it})}{\phi} \tag{3.2}$$

(Diggle,Liang und Zeger, 1986).
Da im generalisierten linearen Modell vorausgesetzt wird, dass die Beobachtungen unabhängig sind ist es nicht zur Analyse von longitudinalen Daten geeignet. Liang und Zeger versuchen im weiteren Verlauf das generalisierte lineare Modell so zu erweitern, dass es auch für korrelierte Daten gültig ist.

4 Schätzgleichungen

4.1 Unabhängige Schätzgleichungen

Im folgendem stellen Liang und Zeger einen Schätzer $\hat{\beta}_I$ von β vor, der sich unter der Arbeitshypothese ergibt, dass die wiederholten Beobachtungen von einem Individuum unabhängig voneinander sind. Unter der arbeitenden Unabhängigkeitsannahme hat die Zielgleichung von einer Likelihood Analyse die Form

$$U_I(\beta) = \sum_{i}^{K} X_i^T \Delta_i S_i = 0, \tag{4.1}$$

wobei $\Delta_i = diag(\frac{\partial \theta_{it}}{\partial \eta_{it}})$ eine $n \times n$ Matrix und $S_i = Y_i - a'(\theta)$ von der Ordnung $n \times 1$ für das i-te Subjekt ist (Diggle,Liang und Zeger, 1986).
Definiert man nun für jedes i eine $n \times n$ Diagonalmatrix $A_i = diag\{a''(\theta_{it})\}$, erhält man unter milden Regularitätsbedingungen folgendes Theorem.

Theorem 1. *Der Schätzer $\hat{\beta}_I$ von β ist konsistent und $K^{1/2}(\hat{\beta}_I - \beta)$ ist asymptotisch multivariat normalverteilt, wenn $K \to \infty$ mit Mittelwert 0 und der Kovarianzmatrix V_1 gegeben durch*

$$\begin{aligned} V_I = \lim_{K\to\infty} K\left(\sum_{i=1}^{K} X_i^T \Delta_i A_i \Delta_i X_i\right)^{-1} \left(\sum_{i=1}^{K} X_i^T \Delta_i cov(Y_i) \Delta_i X_i\right) \left(\sum_{i=1}^{K} X_i^T \Delta_i A_i \Delta_i X_i\right)^{-1} \\ = \lim_{K\to\infty} K\big\{H_1(\beta)\big\}^{-1} H_2(\beta) \big\{H_1(\beta)\big\}^{-1} \end{aligned} \tag{4.2}$$

wo die Momentberechnung für die $Y_i's$ in Bezug auf das wahre zugrundeleigende Modell genommen wurden.

Die Schätzung von ϕ ist nicht nötig um V_I zu schätzen, obwohl letztere sogar eine Funktion von ϕ ist (Diggle, Liang und Zeger, 1986).
Der Schätzer $\hat{\beta}_I$ bietet den Vorteil, dass sowohl $\hat{\beta}_I$, als auch var($\hat{\beta}_I$) konsistent sind, wenn lediglich die Regression korrekt spezifiziert wurde, was das Hauptinteresse ist. Dies erfordert, dass fehlende Werte komplett zufällig fehlen. Der Hauptnachteil von dem Schätzer $\hat{\beta}_I$ ist, dass er in Fällen mit hoher Autokorrelation möglicherweise nicht sehr Effizient ist (Diggle, Liang und Zeger, 1986).

4.2 Generalisierte Schätzgleichungen

Im folgenden stellen Liang und Zeger eine Klasse von generalisierte Schätzgleichung vor, welche die Korrelationen berücksichtigen um somit die Effizienz zu erhöhen. Die attraktivsten Eigenschaften der GEE Methodologie sind : Allgemeinheit - stetige, diskrete und binäre Messgrößen werden in gleicherweise wie in klassischen generalisierten linearen Modellen gehandhabt; Einfachheit - viele Analysen können mit herkömmlichen Regressions-Paketen ausgeführt werden. Liang und Zeger stellten fest, dass es im weiteren wichtig ist zwischen Populations-Durchschnitt und Individuen-spezifischen Regressionsparametern zu unterschieden. Die resultierenden Schätzer von β bleiben konsistent. Zudem sind konsistente Varianzschätzer vorhanden, wenn die schwache Annahme erfüllt ist, dass ein gewichteter Durchschnitt der geschätzten Korrelationsmatrix gegen eine feste Matrix konvergiert (Diggle, Liang und Zeger, 1986).
Sei nun $R(\alpha)$ eine symmetrische $n \times n$ Matrix, welche alle Anforderungen einer Korrelationsmatrix erfüllt und sei α ein $s \times 1$ Vektor, welcher komplett $R(\alpha)$ bestimmt. Wobei $R(\alpha)$ eine arbeitende Korrelationsmatrix ist. Definiere

$$V_i = \frac{A_i^{1/2} R(\alpha) A_i^{1/2}}{\phi}, \tag{4.3}$$

welches der $cov(Y_i)$ entspricht, wenn $R(\alpha)$ tatsächlich die wahre Korrelationsmatrix der $Y_i's$ ist (Diggle, Liang und Zeger, 1986).
Liang und Zeger definieren nun die generalisierten Schätzgleichungen als

$$\sum_{i=1}^{K} D_i^T V_i^{-1} S_i = 0, \tag{4.4}$$

wobei $D_i = d\{a_i'(\theta)\}/d\beta = A_i \delta_i X_i$ ist. Die Gleichung (4.2) wird zur Unabhängigkeitsgleichung (3.1), wenn man $R(\alpha)$ als die Identitätsmatrix bestimmt.
Für jedes i , ist $U_i(\beta, \alpha) = D_i^T V_i^{-1} S_i$ gleich der Funktion von dem von Wedderburn

(1974) und McCullagh (1983) befürworteten Quasi-Likelihood Ansatz, außer das die $V_i's$ hier nicht nur eine Funktion von β, sondern auch von α sind. Gleichung (4.2) kann ausgedrückt werden, als eine Funktion nur von β , indem man das α in Gleichung (4.3) und (4.4) durch $\hat{\alpha}(y, \beta, \phi)$. $\hat{\alpha}$ ist ein $K^{1/2}$-konsistenter Schätzer von α, wenn β und ϕ bekannt sind und es gilt, dass $K^{1/2}(\hat{\alpha} - \alpha) = O_p(1)$.
Wenn man jetzt noch ϕ durch $\hat{\phi}(Y, \beta)$, ein $K^{1/2}$-konsistenter Schätzer wenn β bekannt ist, ersetzt hat die Gleichung (4.4) nun die Form

$$\sum_{i=1}^{K} U_i[\beta, \hat{\alpha}\{\beta, \hat{\phi}(\beta)\}] = 0. \tag{4.5}$$

Die Lösung der Gleichung ist definiert, als $\hat{\beta}_G$. Das nächshte Theorem besagt die große Stichproben Eigenschaft von $\hat{\beta}_G$.

Theorem 2. *Unter milden Regularitätsbedingungen und gegeben, dass*

1. *$\hat{\alpha}$ ist $K^{1/2}$-konsistent gegeben β und ϕ ;*
2. *$\hat{\phi}$ ist $K^{1/2}$-konsistent gegeben β; und*
3. *$|\partial\hat{\alpha}(\beta, \phi)/\partial\phi| \leq H(Y, \beta)$, welches $O_p(1)$ ist;*

dann ist $K^{1/2}(\hat{\beta}_G - \beta)$ asymptotisch multivariat normalverteilt mit Mittelwert 0 und der Kovarianzmatrix V_G, gegeben durch

$$V_G = \lim_{K\to\infty} K\Big(\sum_{i=1}^{K} D_i^T V_i^{-1} D_i\Big)^{-1} \Big\{\sum_{i=1}^{K} D_i^T V_i^{-1} cov(Y_i) V_i^{-1} D_i\Big\} \Big(\sum_{i=1}^{K} D_i^T V_i^{-1} D_i\Big)^{-1}$$

Man erhält die Varianzschätzung $\hat{V}_G$ von $\hat{\beta}_G$, indem man $cov(Y_i)$ durch $S_i S_i^T$ und β, ϕ, α durch ihre Schätzer in dem Ausdruck V_G ersetzt (Diggle,Liang und Zeger, 1986).
Wie in dem Fall der Unabhängigkeit hängt die Konsistenz von $\hat{\beta}_G$ und $\hat{V}_G$ nur von der richtigen Spezifizierung des Mittelwertes und nicht von der korrekten Wahl von R ab. Dies erfordert wieder, dass fehlende Beobachtungen komplett zufällig fehlen (Rubin). Des weiteren hängt die asymptotische Varianz von $\hat{\beta}_G$ nicht von der Wahl der Schätzer für α und ϕ, unter denen die $K^{1/2}$-konsistent sind ab (Diggle,Liang und Zeger, 1986).
Vergleichbare Ergebnisse sind aus Fällen mit Normalverteilen Daten und der Quasi-Likelihood bekannt, wo die Varianz der Regressionsparameter nicht von der Wahl des Schätzers von ϕ abhängt. Das Problem von Liang und Zeger ist, dass sich die Ergebnisse von der Wahl der Schätzgleichungen für β abhängen, wo der Likelihood

nicht komplett bestimmt ist. Hier ist der individuelle Beitrag ,U_i, ein Produkt aus zwei Ausdrücken. Der erste umfasst α aber nicht die Daten und der zweite ist unabhängig von α mit Erwartungswert Null. Dann ist $\sum E(\hat{c}U_{ij}\partial\alpha)$ gleich $o_p(K)$ und die $var(\hat{\beta}_G)$ hängt nicht mehr von $\hat{\alpha}$ und $\hat{\phi}$ ab (Diggle,Liang und Zeger, 1986).

Diese Schätzgleichungen, jetzt weitestgehend als "GEE"bekannt ist nicht nur für die Analyse von longitudinalen Daten sondern auch für andere Formen Cluster Daten geeignet (Diggle,Liang und Zeger, 1986).

4.2.1 Verbindung zur der Gauss-Newton Methode

Um $\hat{\beta}_G$ zu berechnen, wiederholen Liang und Zeger einer modifiziertes Fisher Scoring für β und einer Momentschätzung von α und ϕ. Gegeben geltender Schätzung $\hat{\alpha}$ und $\hat{\phi}$ von den Störgrößen, schlagen Liang und Zeger folgendes modifiziertes Verfahren für β vor:

$$\hat{\beta}_{j+1} = \hat{\beta}_j - \left\{ \sum_i^K D_i^T(\hat{\beta}_j)\tilde{V}_i^{-1}(\hat{\beta}_j)D_j(\hat{\beta}_j) \right\}^{-1} \left\{ \sum_i^K D_i^T(\hat{\beta}_j)\tilde{V}_i^{-1}(\hat{\beta}_j)S_i(\hat{\beta}_j) \right\}. \quad (4.6)$$

mit $\tilde{V}(\beta) = V_i[\beta, \hat{\alpha}\{\beta, \hat{\phi}(\beta)\}]$. Dieses Verfahren kann laut Liang und Zeger als eine Modifikation von Fisher's Scoring Methode gesehen werden, in dem der Grenzwert von dem Erwartungswert von der Ableitung von $\sum U_i[\beta, \hat{\alpha}\{\beta, \hat{\phi}(\beta)\}]$ zur Korrektion genutzt wird.
Definiert man nun $D = (D_1^T, \ldots, D_K^T)^T$, $S = (S_1^T, \ldots, S_K^T)^T$ und sei $\tilde{V}$ eine $nK \times nK$ Diagonalmatrix mit den $\tilde{V}_i$'s als Diagonalelementen. Definiere die modifizierte abhängige Variable

$$Z = D\beta - S,$$

dann ist das iterative Verfahren (4.4) zur Berechnung von $\hat{\beta}_G$ äquivalent zur Durchführung einer iterativen neu-gewichteten linearen Regression von Z auf D mit Gewichtung $\tilde{V}^{-1}$(Diggle,Liang und Zeger, 1986) .

4.2.2 Die Schätzer von α und ϕ

Bei einer gegebenen Iteration kann der Korrelationsparameter α und der Skalaparameter ϕ durch die bestehenden Pearson Residuen geschätzt werden. Die Pearson

Residuen sind definiert durch

$$\hat{r}_{it} = \{y_{it} - a'(\hat{\theta}_{it})\}/\{a''(\hat{\theta}_i t)\}^{1/2},$$

wobei $\hat{\theta}_{it}$ von dem bestehenden Wert von β abhängt. Man kann nun ϕ schätzen, durch

$$\hat{\phi}^{-1} = \sum_{i=1}^{K} \sum_{t=1}^{n_i} \hat{r}_{it}^2/(N-p).$$

mit $N = \sum n_i$. Das ist das longitudinale Analogon zu der Pearson Statistik (Diggle,Liang und Zeger, 1986). Um α konsistent zu schätzen, bedienen sich Liang und Zeger der Stärke über die K Individuen. Der spezifische Schätzer hängt von der Wahl von $R(\alpha)$ ab. Der allgemeine Ansatz ist es α, mittels folgender Funktion zu schätzen

$$\hat{R}_{uv} = \sum_{i=1}^{K} \hat{r}_{iu}\hat{r}_{iv}/(N-p)$$

Alternative Schätzer für ϕ, wie solche die auf der Log-Likelihood basieren, sind vorhanden. Jedoch sind die analogen Schätzer für α nicht vorhanden, da Liang und Zeger nicht des gesamte gemeinsame Verteilung von Y_i bestimmen.

4.2.3 Beispiele von Schätzgleichungen

In diesem Abschnitt werden drei verschiedene spezfische Wahlen von $R(\alpha)$ angenommen, wobei die Anzahl der Störparameter und der Schätzer von α von Fall zu Fall variieren.

Beispiel 1. Sei $\alpha = (\alpha_1 \ldots, \alpha_{n-1})$, wobei $\alpha_t = corr(Y_{it,}, Y_{i,t+1})$ für $t = 1, \ldots, n-1$ ist. Ein natürlicher Schätz von α_t, bei gegebenen β und ϕ ist :

$$\hat{\alpha}_t = \phi \sum_{i=1}^{K} \hat{r}_{it}\hat{r}_{i,t+1}/(K-p).$$

Sei $R(\alpha)$ nun tridiagonal mit $R_{t,t+1} = \alpha_t$. Dann ist dies äquivalent zu den einmalig Abhängigen Modell. Eine Schätzung von ϕ ist zur Berechnung von $\hat{\beta}_G$ und $\hat{V}_G$ nicht nötig, da sie sich in der Berechnung von V_i rauskürzen. In dem Spezialfall,
wo s = 1 und $\alpha_t = \alpha(t = 1, \ldots, n-1)$ gilt, kann das allgemeine α geschätzt werden

durch (Diggle, Liang und Zeger, 1986):

$$\hat{\alpha} = \sum_{t=1}^{n-1} \hat{\alpha}_t/(n-1).$$

Beispiel 2. Sei s = 1 und angenommen dass $corr(y_{it}, y_{it'}) = \alpha$ für alle $t \neq t'$. Dies ist die austauschbare Korrelationsstruktur, welche man für ein zufälliges Effekt-Modell mit einen zufälligen Level für jedes Subjekt erhält.
Wenn ϕ gegeben ist, kann α geschätzt werden durch:

$$\hat{\alpha} = \phi \sum_{i=1}^{K} \sum_{t>t'} \hat{r}_{it}\hat{r}_{it'} \Big/ \Big\{ \sum_{i=1}^{K} \frac{1}{2} n_i(n_i-1) - p \Big\}.$$

Die Schätzung von ϕ ist nicht nötig um $\hat{\beta}_G$ und $\hat{V}_G$ zu erhalten. Mit dieser Annahme ist eine willkürliche Anzahl von Beobachtungen und Beobachtungszeitpunkten für jedes Individuum möglich (Diggle ,Liang und Zeger, 1986).

Beispiel 3. Sei $corr(y_{it}, y_{it'}) = \alpha^{|t-t'|}$. Wenn y_{it} normalverteilt ist, ist die Korrelationsstruktur von der stetigen Zeit analoge zu der Autokorrelaton erster Ordnung, AR -1. Da nach diesem Modell $E(\hat{r}_{it}\hat{r}_{it'}) \doteq \alpha^{|t-t'|}$ kann α durch die Steigung von einer Regression $log(\hat{r}_{it}\hat{r}_{it'})$ auf $log(|t-t'|)$ geschätzt werden. In diesem arbeitenden Modell können beliebige Anzahlen und Abstände von Beobachtungen verwendet werden. Jedoch muss ϕ zur Bestimmung von $\hat{\beta}_G$ und $\hat{V}_G$ berechnet werden (Diggle, Liang und Zeger, 1986).

4.3 Effizienzbetrachtung der Schätzer

Im folgenden werden zwei einfache Datenstrukturen betrachtet, um folgende zwei Fragen zu beantworten. Das Interesse liegt zum einem in der Beantwortung der Frage, wie viel effizienter $\hat{\beta}_G$ als $\hat{\beta}_I$ ist. Zum anderen, wie sich $\hat{\beta}_G$ und $\hat{\beta}_I$ mit dem Maximum Likelihood Schätzer vergleichen lassen, wenn zusätzliche Verteilungsannahmen über die Y_i's getroffen werden (Diggle, Liang und Zeger, 1986).
Um die erste Frage zu beantworten, betrachtet man ein generalisiertes lineares Modell mit natürlichem Link, so dass

$$\theta_{it} = x_{it}\beta \qquad (t = 1, \ldots, 10).$$

Liang und Zeger nehmen im folgendem an, dass jedes $X_i = (x_{i1}, \ldots, x_{i,10})'$ von einer Verteilung mit Mittelwert $(0.1, 0.2, \ldots, 1.0)'$ und endlicher Kovarianz generiert wurde. Tabele 1 zeigt die asymptotische relative Effizienz der $\hat{\beta}_I$ und $\hat{\beta}_G$'s für drei verschiedene Korrelationsannahmen für den generalisierten Schätzer bei dem die Korrelationsmatrix korrekt bestimmt wurde. Die Korrelationsstrukturen sind önedependent"(einmal-abhängig), ëxchangeable"(austauschbar) und "first-order autoregressiv"(Autoregressiv erster Ordnung). Entspricht bsp. 2,3 und 4. Die oberen und unteren Einträge für α sind jeweils 0.3 und 0.7.
Es sind kleine Unterschiede zwischen den $\hat{\beta}_I$ und $\hat{\beta}_G$'s erkennbar, wenn die wahre Korrelation mäßig ist, zum Beispiel 0.3. Jedoch zeigen die unteren Einträge von Spalte 1, dass wesentliche Verbesserungen gemacht werden können, wenn α groß ist, indem die Korrelationsmatrix korrekt bestimmt wird. Die Effizienz von $\hat{\beta}_I$ bezogen auf die $\hat{\beta}_G$, unter Verwendung der richtigen Korrelationsmatrix ist am niedrigsten, mit 0.74, wenn R die einmal-abhängige Form aufweist und am höchsten, mit 0.99, wenn R das austauschbare Muster hat. Das im obrigen Fall $\hat{\beta}_I$ bezogen auf $\hat{\beta}_G$ effizient ist liegt daran, dass $n_i = 10$ ist für alle i, sodass die zusätzliche binomische Veränderung, welche von der austauschbaren Korrelation eingeführt wurde für alle Subjekte die gleiche ist und keine Fehlgewichtungen auftauchen, wenn man sie ignoriert. Angenommen das n_i Werte von 1 bis 8 mit gleichen Wahrscheinlichkeiten annimmt, würde die relative Effizienz von $\hat{\beta}_I$ auf 0.82 fallen. Die Ergebnisse aus Tabelle 1 ungeachtet von der zugrundeliegenden marginalen Verteilung (Diggle, Liang und Zeger, 1986).

Tabelle 1. Asymptotische relative Effizienz von $\hat{\beta}_I$ und $\hat{\beta}_G$ auf den generalisierten Schätzer mit richtig bestimmter Korrelationsmatrix für $\eta = \beta_0 + \beta_1 t/10$. Hier, $\beta_0 = \beta_1 = 1$,$n_i = 10$. Für die oberen Einträge ist $\alpha = 0.3$ und für die unteren $\alpha = 0.7$.

	Working R			
True R	Independence	1-Dependence	Exchangeable	AR-1
1-Dependence	0.97	1.0	0.97	0.99
	0.74	1.0	0.74	0.81
Exchangeable	0.99	0.95	1.0	0.95
	0.99	0.23	1.0	0.72
AR-1	0.97	0.99	0.97	1.0
	0.88	0.75	0.88	1.0

Um die zweite Frage zu beantworten, betrachten Liang und Zeger eine zwei Stichproben Konfiguration mit binären Ergebnissen. Individuen sind in zwei Gruppen, mit Randerwartung logit $\{E(y_{it})\} = \beta_0 + \beta_1 x_i$, mit $x_i = 0$ für Gruppe 0 und 1 für Gruppe 1. Es wird angenommen, dass die wiederholten Beobachtungen von ein Markov Kette erster Ordnung stammen, mit α als Autokorrelation erster Verzögerung (Diggle,Liang und Zeger, 1986).
In Tabelle 2 sieht man den Vergleich von der asymptotischen relativen Effizienz von $\hat{\beta}_I$ und $\hat{\beta}_G$ unter Verwendung der AR-1 Korrelationsstruktur. bsp 4. Für die oberen Einträge gilt $n_i = 10$ für alle i; für die unteren, $n_i = 1$ bis 8 mit gleichen Wahrscheinlichkeiten. Die Ergebnisse zeigen, dass sowohl $\hat{\beta}_I$ als auch $\hat{\beta}_G$ hoch effizient sind für kleines α. Wenn sich das α erhöht, bleibt $\hat{\beta}_G$ im Gegensatz zu $\hat{\beta}_I$ nahezu vollständig effizient. Der Kontrast zwischen $\hat{\beta}_I$ und $\hat{\beta}_G$ wird sogar noch stärker wenn sich die Stichprobenumfänge unterscheiden (Diggle, Liang und Zeger, 1986).

Tabelle 2. Asymptotische relative Effizienz von $\hat{\beta}_I$ und $\hat{\beta}_G$ mit angenommener AR-1 Korrelationsstruktur für den Maximum Likelihood Schätzer für eine Markov Kette erster Ordnung mit $\theta_{it} = \beta_0 + \beta_1 x_i$. Wobei $x_i = 0$ für Gruppe 0, $x_1 = 1$ für Gruppe 1. Hier, $\beta_0 = 0$, $\beta_1 = 1$, und für die oberen Einträge $n_i = 10$, für die unteren Einträge $n_i = 1, \ldots, 8$ mit gleichen Wahrscheinlichkeiten.

	Correlation α						
	0.0	0.1	0.2	0.3	0.5	0.7	0.9
$\hat{\beta}_I$	1.0	1.0	0.99	0.97	0.94	0.91	0.92
	1.0	1.0	0.98	0.96	0.92	0.86	0.81
$\hat{\beta}_G(AR{-}1)$	1.0	1.0	0.99	0.99	0.98	0.97	0.98
	1.0	1.0	0.99	0.99	0.98	0.98	0.99

5 Diskussion

Die hergeleiteten Schätzfunktionen sind für longitudinale Daten gültig. Es exisitieren jedoch kleine Nachteile. Ein Nachteil ist dass die robusten Eigenschaften nur gültig sind, wenn die Anzahl der Subjekte gegen unendlich geht. Desweiteren sind die geschätzten Standardfehler nur zuverlässig, wenn die Daten aus vielen Kleinen Datensätzen bestehen. Ein weiterer möglicher Nachteil der generalisierten Schätzgleichungen ist, dass diese nicht mit fehlenden Werten umgehen können, außer diese fehlen komplett zufällig. Eine Erweiterung der GEE von Robins, Rotnitzky und Zhao (1995) ist auch unter schwächeren Annahmen über komplett zufällig fehlende Daten gültig. Wenn R die wahre Korrelation ist, wird die Annahme, dass die Daten komplett zufällig fehlen unnötig (Diggle, Liang und Zeger, 1986).

Literaturverzeichnis

Diggle, Liang und Zeger (1986). *Longitudinal Data Analysis Using Generalized Linear Models* . Biometrika, Vol. 73.

Prof. Dr. Anne-Laure Boulesteix (2012). *Volesungskript: Analyse Longitudinaler*

Prof. Dr. Helmut Küchenhoff (2014). *Volesungskript: Generalisierte Regression*

Claudia Funk-Hüsges (2002). *Generalisierte Lineare Modelle mit zufälligen Effekten und variierenden Koeffizienten*

Erklärung zur Urheberschaft

Hiermit versichere ich, dass ich die vorliegende Seminararbeit selbstständig aus den Vorlagen von Andrea Wiencierz angefertigt habe.

München, den 30. Mai 2014

(Markus Riedl)